THE PARAPSYCHOLOGICAL ASSOCIATION, INC.

47TH ANNUAL CONVENTION

AUGUST, 5-8, 2004

PROCEEDINGS OF PRESENTED PAPERS

Vienna University

Stefan Schmidt
Program Chair

THE PARAPSYCHOLOGICAL ASSOCIATION, INC.
47TH ANNUAL CONVENTION

AUGUST, 5-8, 2004
VIENNA UNIVERSITY

PROGRAM CHAIR
Stefan Schmidt

ARRANGEMENTS COMMITTEE
Peter Mulacz
Manfred Kremser
Gerhard Holischka

PRODUCTION OF PROCEEDINGS
Andreas Sommer

PROGRAM COMMITTEE

Cheryl Alexander
Carlos Alvarado
Eberhard Bauer
Dick Bierman
Holger Bösch
Richard Broughton
Donald Burdick
Kathy Dalton
Hoyt Edge
Christopher French
Erlendur Haraldsson
Gerd Hövelmann
Harvey Irwin
Edwin May
Robert Morris
Roger Nelson
John Palmer
Dean Radin
Chris Roe
Hartmann Römer
Christine Simmonds
James Spottiswoode
Paul Stevens
Nikolaus v. Stillfried
Jiri Wackermann
Joakim Westerlund

INTRODUCTION

This year's Parapsychological Association Convention is held in Vienna, Austria. It is hosted by the Austrian Society for Parapsychology, which had its 75th anniversary two years ago. We are looking forward to an exciting and busy program with 50 different contributions from 14 countries. This year, 72 percent of the accepted papers, posters and research briefs are from European countries, 20 percent from the US and six percent from Asian Countries. Seventeen contributions alone are from the UK were Psi research is meanwhile taking place at five universities.

The fifty contributions presented here were selected from all submissions in a lively review process by our excellent program committee. All submissions were provided with a constructive and informative feedback. Thus the review process did not only serve to accept or reject manuscripts but also helped in many cases to improve the quality of the contributions through the input of the peers. This is due to the hard work of the program committee and my thanks go to: Cheryl Alexander, Carlos Alvarado, Eberhard Bauer, Dick Bierman, Holger Bösch, Richard Broughton, Donald Burdick, Kathy Dalton, Hoyt Edge, Christopher French, Erlendur Haraldsson, Gerd Hövelmann, Harvey Irwin, Ed May, Robert Morris, Roger Nelson, John Palmer, Dean Radin, Chris Roe, Hartmann Römer, Christine Simmonds, James Spottiswoode, Paul Stevens, Niko v. Stillfried, Jiri Wackermann and Joakim Westerlund.

My thanks also go to Peter Mulacz who as Arrangement Chair did an excellent work to make this conference possible and always responded helpfully to my requests for support.

This year's proceedings have a slightly new look. This is due to a new submission procedure where authors wrote their manuscripts directly into a template to guarantee a somewhat homogenous appearance. The design of this template as well as the laborious procedure of compiling and formatting all manuscripts was done by Andreas Sommer. He was furthermore a helpful assistant in many difficult situations and I am grateful to him.

Finally I want to thank Harald Walach and the Samueli Institute for providing me with the necessary time and resources to organise this year's program.

Stefan Schmidt
University Hospital Freiburg
Samueli Institute, Europe

TABLE OF CONTENTS

Floor Presentations

Bösch, Holger	*Reanalyzing a meta-analysis on ESP dating from 1940*	1
Broughton, Richard S.	*Exploring the reliability of the "presentiment" effect*	15
Fox, Jezz & Simmonds, Christine	*Ruling out items in the judging set: a novel protocol for detecting psi in the Ganzfeld*	27
Hinterberger, Thilo; Houtkooper, Joop M. & Kotchoubey, Boris	*Effects of feedback control on slow cortical potentials and random events*	39
Houtkooper, Joop M.	*Exploring volitional strategies in the mind-machine interaction replication*	51
Kittenis, Mario; Caryl, Peter & Stevens, Paul	*Distant psychophysiological interaction effects between related and unrelated participants*	67
Lobach, Eva & Bierman, Dick	*The invisible gaze: three attempts to replicate Sheldrake's staring effects*	77
Lobach, Eva & Bierman, Dick	*Who's calling at this hour? Local sidereal time and telephone telepathy*	91
Lucadou, Walter von & Zahradnik, Frauke	*Predictions of the Model of Pragmatic Information about RSP*	99
May, Edwin C.	*Anticipatory skin conductance responses: a possible example of decision augmentation theory*	113
McClenon, James	*What is to be done: evaluating the Ritual Healing Theory*	121
Mulacz, Peter	*Strange people – rare events*	135
Murray, Craig D. & Fox, Jezz	*Body image in respondents with and without prior out-of-body experiences*	145
Pallikari, Fotini	*On the false hypothesis of psi-mediated shift of statistical average in tests with random event generators*	157
Palmer, John	*Synchronicity and psi: how are they related?*	173
Roe, Chris A.; Davey, Russell & Stevens, Paul	*Experimenter effects in laboratory tests of ESP and PK using a common protocol*	185
Roe, Chris A & Holt, Nicola	*A further consideration of the sender as a PK agent in Ganzfeld ESP studies*	197
Roney-Dougal, Serena M. & Solfvin, J.	*Field study of an enhancement effect on lettuce seeds: working in adverse conditions*	207
Savva, Louie; Child, Robert & Smith, Matthew D.	*The precognitive habituation effect: an adaptation using spider stimuli*	223
Sherwood, Simon J.; Roe, Chris A.; Holt, Nicola J. & Wilson Stuart	*Interpersonal psi: exploring the role of the experimenter and the experimental climate in a Ganzfeld telepathy tast*	231
Schmidt, Stefan; Müller, Susanne & Walach, Harald	*Do you know who is on the phone? Replication of an experiment on telephone telepathy*	245
Westerlund, Joakim; Parker, Adrian; Dalkvist, Jan & Goulding, Anneli	*Remarkable correspondences between Ganzfeld mentation and target content – psi or a cognitive illusion?*	255

Westerlund, Joakim & Dalkvist, Jan	*A test of predictions from five studies on telepathic group communication of emotions*	269
Wiseman, Richard & Greening, Emma	*'It's still bending': verbal suggestion and alleged psychokinetic ability*	279
Wiseman, Richard & Watt, Caroline	*Measuring superstitious belief: why lucky charms matter*	291
Wiseman, Richard & O'Keeffe, Ciarán	*Testing alleged mediumship. Methods and results*	299

Panels

Perspectives in the study of mediumship (Chair: Carlos S. Alvarado)

Alvarado, Carlos S.	*On trances and spirit communicators: research ideas from the old literature*	313
Reinsel, Ruth	*Physical mediumship, macro-PK and the autonomic nervous system*	314
Mulacz, Peter	*Physical mediumship: open questions from the past – a challenge for the future*	315
Pilkington, Rosemarie	*The mystery of ectoplasm*	316
Haraldsson, Erlendur	*A survey of mediums in Iceland*	317

Counseling persons with exceptional experiences: the example of IGPP (Chair: Eberhard Bauer)

Bauer, Eberhard	*Introduction: counseling and mental hygiene – the approach of the IGPP*	319
Belz-Merk, Martina	*The role of inconsistency and instability for the origin of exceptional experiences*	321
Fach, Wolfgang	*The phenomenology of exceptional experiences: an analysis of IGPP counseling cases*	322
Kohls, Niko & Walach, Harald	*Exceptional experiences and mental health – results and problems of a questionnaire study*	323

Poster sessions

Auriol, Bernhard M.; Garcia, Frédéric; Puech, Laetitia; et al.	*Agapé: group telepathy. A long-term experimental series*	325
Boldyreva, Liudmila & Sotina, Nina B.	*Classification of phenomena in parapsychology from the point of view of physics. Superfluid vacuum*	347
Ertel, Suitbert	*Psi test feats achieved alone at home: do they disappear under lab control?*	355
Etzold, Eckhard	*Does psi exist and can we prove it? Belief and disbelief in psychokinesis research*	367
Hagel, Johannes & Tschapke, Margot	*The local event detector (LED) – an experimental setup for an exploratory study of correlations between collective emotional events and random number sequences*	379
Hirukawa, Tatsu & Ishikawa, Masato	*Anomalous fluctuation of RNG data in Nebuta summer festival in northeast Japan*	389
McMoneagle, Joseph & May, Edwin C.	*The possible role of intention, attention and expectation in remote viewing*	399

Sartori, Luisa; Massaccesi, Stefano; Martinelli, Massimiliano & Tressoldi, Patrizio	*Physiological correlates of ESP: heart rate differences between targets and non-targets in clairvoyance and precognition forced choice tasks*	407
Shan, Gao	*A primary quantum model of telepathy*	413
Tam, Wai-Cheong Carl & Shiah, Yung-Jong	*Paranormal belief, religiosity and cognitive complexity*	423

Research Briefs

Bem, Daryl J.	*Precognitive avoidance and precognitive déjà vu*	431
Holt, Nicola J.; Delanoy, Deborah L. & Roe, Chris A.	*Creativity, subjective paranormal experiences and altered states of consciousness*	433
Ivtsan, Itai & French, Christopher C.	*Testing the validity of tarot cards: can we distinguish between a real and a control reading?*	437
Krippner, Stanley	*The dreams and visions of Eva Hellström*	441
Joines, William T.; Baumann, Stephen B.; Kim, Jeremy, et al.	*The measurement and characterization of charge accumulation and electromagnetic emission from bioenergy healers*	445
Don, Norman S.; McDonough, Bruce E. & Warren, Charles A.	*Combining the event-related brain potential (ERP) and precognitive habituation techniques: a methodological proposal*	449
Ramakers, Peter	*Response characteristics and lability in an EDA-DMILS study using emotional stimuli*	451
Simmonds, Christine A.	*A qualitative investigation of source monitoring in the context of subjective paranormal experiences*	455
Smith, Matthew D. & Savva, Louie	*Experimenter effects and psi performance using a digital Autoganzfeld system*	461
Wackermann, Jiří; Naranjo Muradás, José Raúl & Pütz, Peter	*Event-related correlations between brain electrical activities of separated human subjects: preliminary results of a replication study*	465
Wilson, Krissy & French, Christopher C.	*Memory conformity and paranormal belief*	469

REANALYZING A META-ANALYSIS ON EXTRA-SENSORY PERCEPTION DATING FROM 1940, THE FIRST COMPREHENSIVE META-ANALYSIS IN THE HISTORY OF SCIENCE

Holger Bösch
Institute of Environmental Medicine and Hospital Epidemiology
University Hospital Freiburg

ABSTRACT

The book *Extra-Sensory Perception After Sixty Years* (*ESP-60*) published in 1940 is the first meta-analysis in book format. *ESP-60* provides a comprehensive and balanced review of 145 reports on ESP experiments published from 1882 to 1939. The authors discuss the quality and reliability of the experimental data and provide an in-depth discussion of arguments questioning the data altogether. They are the first meta-analysts taking into account publication bias, and calculate a 'fail-safe *N*' in order to assure the reliability of their analyses long before such sensitivity analyses became standard practice in meta-analysis.

All statistical analyses provided in *ESP-60* are based on *z*-score statistics. The purpose of this paper is to reanalyze the data using an effect-size based approach. Altogether 139 independent effect-sizes measures could be calculated. The distribution of effect-sizes dramatically deviates from the expected funnel shape around the overall mean $\overline{pi}$ = .57. A strikingly small-study effect was found. An analysis of sample-size quartiles illustrates the highly visible connection between effect-size and sample-size. The smallest studies (Q_1) in the sample (*M* = 36 trials) have an effect-size of $\overline{pi}$ = .99, indicating that on average the hit probability of these trials is 99% where *MCE* is 50%. The largest studies (Q_4) in the sample (M = 113716 trials) have an effect-size of $\overline{pi}$ = .52 indicating that at 52% the hit probability of these trials is just above MCE.

An explanation for the dramatic small-study effect cannot be directly derived from the data. Several factors beside publication bias, such as true heterogeneity, data irregularities and chance must be considered. However, this data is especially susceptible to publication bias because the early experimental approach to ESP evolved from case studies which principally aimed at demonstrating ESP. Although the authors of *ESP-60* have done everything to ensure the completeness of their data, they have underestimated the problem of publication bias.

That studies are published independently of their *p*-value is an ideal that so far, no empirical science has accomplished. Research in ESP is especially affected because ESP studies are generally not very cost and time intensive, factors which are clearly related to publication bias. It is therefore suggested to implement registers for ESP studies and to accept only registered trials for publication. This would allow minimizing the impact of publication bias in future meta-analyses and help to provide a balanced perspective on ESP research altogether.

INTRODUCTION

Extra-Sensory Perception After Sixty Years (*ESP-60*) is a classic in experimental parapsychology and in meta-analysis. Right after its publication the book became an assigned reading at Harvard psychology introductory classes (Broughton, 1991). The outstanding quality and balance of this scholarly book is still evident today. It provides a perfect overview over the first 60 years of experimental research in extra-sensory perception (ESP). Existing shortcomings and challenges of the experimental approach to ESP are discussed in great detail. The book provides answers to the most common criticism regarding ESP research. *ESP-60* combines data from 145 reports with 185 assigned *z*-scores[1] and provides several detailed subgroups analyses. It is a summary of a "wide variety of materials, methods, conditions and results" (Pratt, Rhine, Smith, Stuart & Greenwood, 1940[2], p. 76) including for instance experiments examining the production of local anaesthesia (e.g., Gurney, 1888), experiments examining the precognitive ability to foresee the fall of dice (e.g., Carington, 1935) and experiments examining the feeling of being stared at (Tichener, 1898). However, most experiments have been carried out using a standard 52-card deck of cards or a 25-card deck of cards developed by the perception psychologist Karl Zener especially for ESP experiments (e.g., Rhine, 1934; Pratt, et al., 1940).

Regarding the meta-analytical approach taken, the book remained unmatched in its comprehensiveness until the last two decades of the 20th century, when meta-analysis became a popular tool for research synthesis (e.g. Egger, Smith & O'Rourke, 2001; Rosenthal & DiMatteo, 2001). Pratt and his colleagues already discussed publication bias long before it became one of the most important topics in meta-analysis. They even calculated a 'fail-safe *N*'. In book format it is probably the first meta-analysis altogether.

The data provided by Pratt and his colleagues has not been reanalyzed yet, although some authors made use of parts of the database. Nash (1989). used the database as presented in Table 29 to establish a correlation between the number of hits in the experiments and the respective number of trials and Radin (1997) analyzed a subset of ESP card experiments (Table 7) published from 1934 to 1939 "which acceptably exclude sensory cues" (Pratt et al. , p. 95). Nash was interested in providing evidence for the decline-effect, although his operationalization can be put into question (Bösch & Walach, 2004). Radin focused on the mere magnitude of the effect and believes that no *file drawer* whatsoever is likely to explain the results obtained. He used Rosenthal's fail-safe approach (e.g. 1979) to explore the alleged file drawer and found that it would need 861 studies for every study published "to reduce this body of evidence to a nonsignificant level" (p. 97). However, it is well know that Rosenthal's approach, although widely used, can be very misleading (e.g., Iyengar & Greenhouse, 1988; Scargle, 2000).

The purpose of this paper is to reanalyze the data using an effect-size based approach instead of a *z*-score based approach as used originally. Although the authors have done everything that one can think of to ensure the quality of the data and taken care to safeguard the conclusions drawn, sixty years later new insights can be gained by switching perspective. The paper also reflects on the publication policy of the time. Although Pratt and his colleagues have addressed the problem of publication bias, today the problem is much better understood, not only in its importance but also in its complexity.

METHOD

Sample of Studies

The present analyses are based solely upon data as published in *ESP-60* Table 29. It is the main summary table of the book and is thought to include all available reports on ESP experiments published between 1882 and 1939. The ESP experiments included are experiments were human subjects try to:

> Perceive (or identify) objects (or the experience of another person acting as sender) without the intermediation of the recognized senses. This excludes, regardless of value, all observation of non-experimental character,

[1] This excludes one *z*-score given in parentheses in Table 29 and two *z*-sores given in the addenda of Table 29 which Pratt et al. have "not included in totals" (p. 407). The reasoning, however, could not be pursued.

[2] Everybody discussing or analyzing *ESP-60* in more detail will find that the authorship of the book is a story on its own (e.g., Alvarado, 1986; Keil, 1987). The problem is that although Gaither Pratt is listed as first author on the title page and at the end of the joint preface, which qualifies him as first author, Rhine's name is listed first on the bound cover. This might excuse why some authors quote *ESP-60* with Rhine as the first author. However, sometimes even within a single book, *ESP-60* is quoted differently like e.g. in the Handbook of Parapsychology (Beloff, 1977; Burdick & Kelly, 1977) or in Basic Research in Parapsychology (Rao, 2001; Schlitz & Gruber, 2001). The inconsistency in quoting is widespread and probably also a result of Rhine's lead (Keil, 1987). Rhine regularly quoted *ESP-60* incorrectly (e.g., Rhine, 1942). Even Pratt, as long as associated with the Duke University Parapsychology Laboratory (Keil, 1987) of which Rhine was the head, quoted *ESP-60* with Rhine as first author. Today's most prominent misquotes can be found on the homepage of the Parapsychological Association (2004) and in Rhea White's database Psiline (1991).

> experimental work with animal subjects, all of the tests of ability to *locate* objects (dowsing, etc.), free-association ESP tests with objects (so-called psychometry), and mediumistic experiments (free-association ESP tests with persons as stimuli or foci of attention). (p. 73)

The authors allude that they have made "a reasonable diligent search" (p. 73). They included "failures" which have been "published only through the medium of the newspapers" (p. 73). Some reports, which the authors could not locate, were replaced by reviews and only a very few reports were altogether inaccessible to the authors.

For every ESP experiment reported in Table 29 the respective number of subjects and total number of trials is listed. However, not every experiment is what the authors call a "quantitative trial", i.e. an experiment with an assigned hit probability, an assigned number of trials and an assigned or computable deviation from mean chance expectation (*MCE*), i.e. data from which individual *z*-scores can be calculated. Altogether, Table 29 lists 185 *z*-scores.

However, some entries in Table 29 are not independent. For example, several reports are listed presenting the results of standard card-experiments, with results given on the level of the individual card (1/52) as well as the card value (1/13), the suit of the card (1/4) and the color of the card (1/2) respectively. For all subsequent analyses the dependent entries have been eliminated, i.e. when entries from a single report refer (1) to the same number of trials *and* (2) have hit probabilities suggesting a card experiment, only the entry with the lowest hit probability level respectively, i.e., e.g. 1/52 in standard card experiments, is kept. This procedure reduced the overall study sample from 185 studies to 139 independent studies.

Computation and Analysis of Effect-Sizes

For one-sample multiple-choice data as present in *ESP-60*, Rosenthal & Rubin (1989) suggest an intuitively comprehensible effect-size measure *pi* (proportion index) with *k*, the number of alternative choices available, and *P*, the raw proportion of hits, which will be used here.

$$pi = \frac{P(k-1)}{1 + P(k-2)} \tag{1}$$

The proportion index expresses hit rates of studies with different hit probabilities according to the hit rate of an equally likely two alternative case like e.g. coin flipping (with a fair coin). Thus, if head in a coin flipping experiment wins at a hit rate of 50%, the effect-size pi = .50 indicates that heads and tails came down equally often (*MCE*). In an experiment with a standard 52-card deck of cards, a 1/52 hit rate on a particular card thus also converts to the null value of pi = .50. The range of *pi* like the range of all probability measures is from 0 to 1.

Following Rosenthal & Rubin (1989), the standard error of *pi* ($SE_{(pi)}$) was calculated based on a large-sample normal approximation on the basis of the common values *P* and *pi*, and the total number of trials per experiment, *N*.

$$SE_{(pi)} = \frac{pi(1-pi)}{\sqrt{N * P(1-P)}} \tag{2}$$

In order to combine effect-sizes from different studies a fixed effect model was used, weighted by the inverse of the variance (e.g., Hedges, 1994).

$$\overline{pi} = \frac{\Sigma w_i pi_i}{\Sigma w_i} \tag{3}$$

$$w_i = \frac{1}{SE^2_{(pi)}} \tag{4}$$

Thus large studies as well as very "successful" small studies with high hit rates get large weights, a procedure that gives more weight to effect-sizes that are more reliably estimated.

To determine whether each set of combined *pi* values shared a common effect-size (i.e., was consistent across studies), a χ^2 homogeneity statistic with *m*-1 degrees of freedom was calculated, where *m* is the number of effect-sizes (Rosenthal & Rubin, 1989).

$$\chi^2(m-1) = \Sigma\left(\frac{pi_i - \overline{pi}}{SE_{(pi)_i}}\right)^2 \tag{5}$$

On the basis of the standard error of the combined effect-sizes $SE_{(\overline{pi})}$ a *z*-score statistics was used to determine the statistical significance of the combined effect-sizes (e.g., Hedges, 1994).

$$z = \frac{\overline{pi} - 0.5}{SE_{(\overline{pi})}} \tag{6}$$

$$SE_{(\overline{pi})} = \frac{1}{\sqrt{\Sigma w_i}} \tag{7}$$

Analyses

All the analyses presented here are descriptive by nature. Therefore, statistical significance will generally not be considered in detail. However, because *z*-scores provided the primary unit of analysis in *ESP*-60 and to demonstrate the uniqueness of the sample, the *z*-score distribution of the 139 independent studies will be shown.

Combined effect-sizes. Standardized effect-sizes from individual studies are combined into composite mean weighted effect-sizes ($\overline{pi}$).The respective *SE*, *z*-score and χ^2 homogeneity statistics were calculated. It is assumed that all studies provide estimates for the same overall effect. Therefore a fixed effect model was applied. Heterogeneity among studies was not investigated in any detail because it is considered to be primarily the result of publication bias. As a secondary analysis and to examine the impact of the weighting procedure, non-weighted composite mean effect-sizes (*PI*) were calculated.

Funnel plot. The funnel plot is used as the main tool to explore the effect-size distribution of studies. Two approaches have been used to examine the hypothesis that the effect-size distribution in the funnel plot is symmetrical, i.e. to test the hypothesis that the effect-size is independent of sample-size. First, and on the basis of Begg & Mazumdar's (1994) approach, a rank correlation between effect-size and sample-size was performed. Second, a regression analysis with sample-size (*ln*) as predictor and effect-size as independent variable was carried out (Egger, Smith, Schneider & Minder, 1997). To further illustrate the effect-size sample-size dependency, the sample was split into quartiles of sample-size.

RESULTS

As can be seen in Figure 1, the first culmination of ESP experiments dates back to the late 1880s. It is a result of the foundation of the Society for Psychical Research in 1882, which was the starting point of the scientific approach to ESP altogether. The second culmination occurred 50 years later in the late 1930s and is primarily a result of the efforts of Rhine and his colleagues working at Duke University.

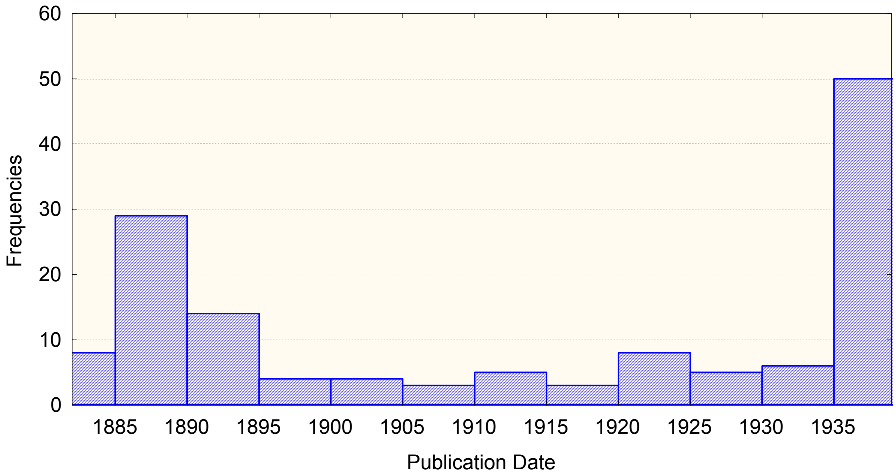


Fig. 1: Distribution of all 139 independent studies gathered in *ESP-60* in time (1882-1938)

The *z*-scores listed in *ESP-60* are stunning. Of the 139 independent *z*-scores, 38 are greater than 10 (see Figure 2). This means that 38 *z*-scores reported in *ESP-60* are more than 10 standard deviations away from *MCE*. The distribution of *z*-scores is very heterogeneous and clearly not normally distributed. From the 41 *z*-scores smaller than 1.96, i.e. statistically not significant, 14 are negative.

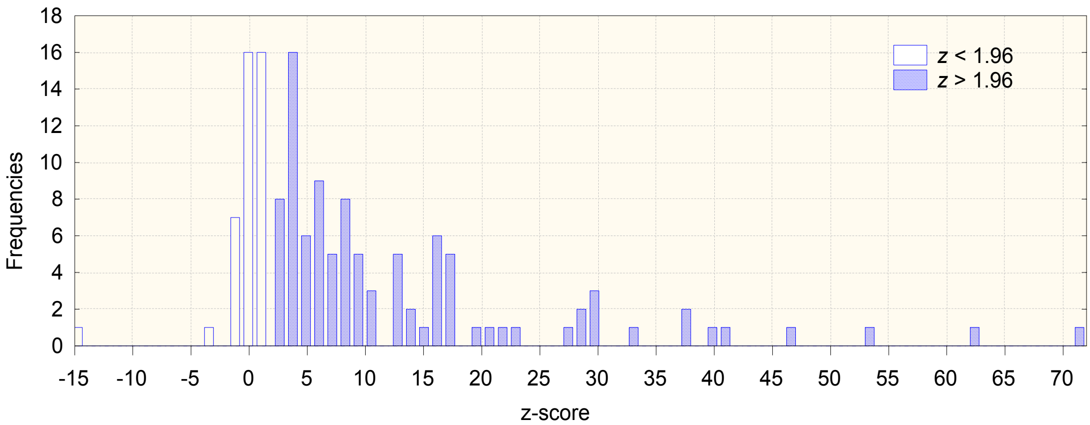


Fig. 2: *Z*-score distribution of the 139 independent *ESP-60* studies.

The main finding of this study is the striking asymmetry in the funnel plot (see Figure 3). The asymmetric distribution of effect-sizes is outstanding in two respects. First, the larger studies do not distribute around the overall mean effect-size of $\overline{pi}$ = .57 (see dotted line on Figure 3). Second, in comparison with the majority of the smaller studies which are distributed to the right of the mean, only a very few of the smaller studies can be found to the left of the mean. Consequently, the whole distribution of effect-sizes is not even close to the theoretically expected symmetrical funnel shape.

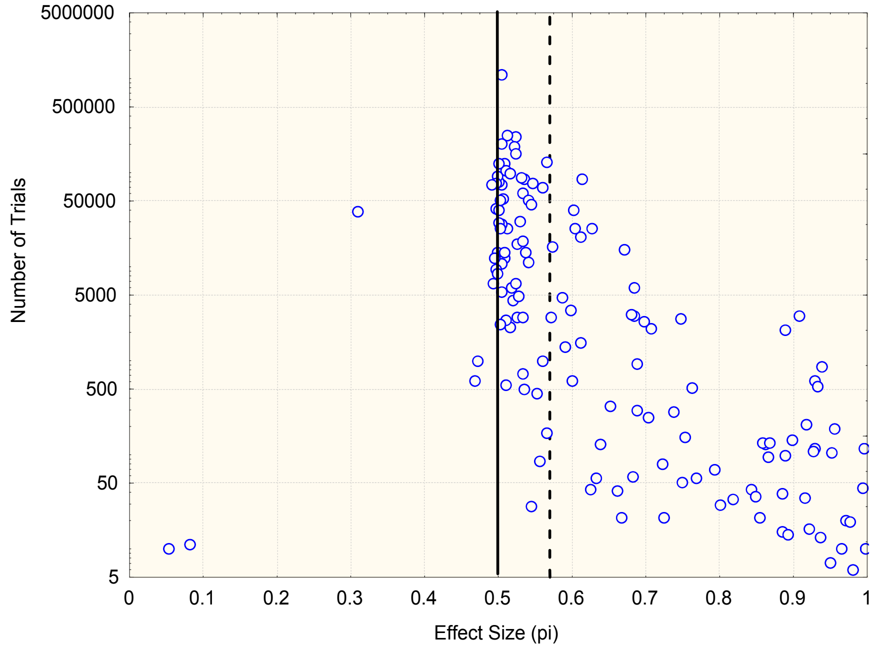


Fig. 3: Funnel plot of all 139 independent *ESP-60* studies.

The highly significant correlation between effect-size and sample-size (r_s = -.67, p = 1.11*10^{-19}) as well as the highly significant regression equation with sample-size (*ln*) as predictor and effect-size as independent variable (R^2 = .37, $F(1/137)$ = 80.22, p = 2.25*10^{-15}) confirm the result of the visual inspection of the funnel plot, i.e. an asymmetric distribution of effect-sizes. Table 1 illustrates the dependency between sample-size and effect-size.

Tabel 1: Summary statistics of all independent *ESP-60* studies and their quartiles.

	n	*M* trials	$\overline{pi}$	*SE*	*Z*	*Min* trials	*Max* trials	χ^2	*M* py
Overall	139	31736	.57	0.00026	272.53*	6	1104100	327112*	1913
Sample-size									
Q_1	34	36	.99	0.00183	267.42*	6	96	275*	1899
Q_2	35	611	.97	0.00087	546.56*	106	2300	7011*	1901
Q_3	34	8672	.56	0.00115	54.79*	2464	24950	7228*	1916
Q_4	36	113716	.52	0.00028	65.08*	25000	1104100	9389*	1937

Note. py = publication year
* p < .001

The effect-size $\overline{pi}$ = .99 of the 34 smallest studies (Q_1) indicates that almost every trial in these experiments was a hit. At 97% the hit rate for the 2nd quartile is of the same magnitude. The overall *z*-scores for the 1st and 2nd quartile are enormous. Nevertheless, a closer look at the *z*-score distribution on study level shows that 22 of the altogether 38 *z*-scores greater than 10 are not from the lower but from the upper two quartiles. On the level of individual studies, *z*-scores and sample-size are not correlated to one another (r = .05, p = .57).

The decline of effect-size from one quartile to the next ends with an effect-size of $\overline{pi}$ = .52 in the 4th quartile. Although the effect-size is highly statistically significant in all quartiles, the phenomenology of ESP in the quartiles is quite different. Smaller sample experiments produce hits almost on demand, with subjects rarely to "fail" suggesting ESP to be an almost omnipotent ability. Larger sample experiments on the other hand reduce ESP to a statistically measurable anomaly.

The funnel plot (see Figure 3) suggests that the overall effect-sizes of Q_1 and Q_2 as shown in Table 1 are disproportionally large. However, this is a result of the weighing procedure, which gives large weights to small studies with high hit rates. Two studies stand out. In Q_1 Lombroso (1890) who published a study with 10 trials ($\overline{pi}$ = .99) and in Q_2 Shield (1887) who published a study with 118 trials ($\overline{pi}$ = .99). However, excluding these studies from the analyses only marginally changes the effect-size of Q_1 (n = 33, $\overline{pi}$ = .98), but the effect-size of Q_2 (n = 34, $\overline{pi}$ = .86) and the effect-size of the overall sample (n = 137, $\overline{pi}$ = .53) drop considerably.

The non-weighted mean effect-sizes also demonstrate the impact of the weighting procedure. The non-weighted overall mean effect-size (*PI* = .65) is bigger than the weighted overall mean. The two lower quartiles Q_1 (*PI* = .78) and Q_2 (*PI* = .73) are considerably different from the respective weighted means, whereas the two upper quartiles Q_3 (*PI* = .57) and Q_4 (*PI* = .52) are almost not affected. However, the conspicuous pattern, i.e. the decline of effect-size from on quartile to the next remains intact. The pattern clearly is not the result of weighting.

As can be seen in Table 1, time (publication year) is a confounding variable of sample-size. The first two quartiles predominantly comprise of studies from the 19th century. In the subsequent quartiles studies from this early period are clearly underrepresented. The 4th quartile comprises only of studies from 1928 to 1939.

Until 1920, i.e. when half the studies in the sample (*n* = 70) were carried out, sample-size is not a function of time ($r = -.04$, $p = .74$). During that time the sample-size, although clearly not normally distributed, remained stable on average (M = 1587, *Mdn* = 123). After 1920 the sample-size of studies (*n* = 69) increased continuously ($r = .25$, $p = .04$). The studies in this period of time are clearly larger (M = 62321, *Mdn* = 25000).

Although the focus in this paper is on the connection between effect-size and sample-size, it should be noted that there is an almost equally strong connection between effect-size and time. The effect-size continuously declines over the years ($r = -.47$, $p = 5.43 \cdot 10^{-9}$). Whether this is the result of improving study quality or larger sample-sizes or another factor e.g. a unique psychological of physical time effect, is impossible to tell. However, because researchers have for a very long time almost exclusively been concerned with extraordinary claims and extraordinary results, far beyond the need of statistics, the increase of sample-size of the average experiment indicates a change in the research paradigm.

It is quite evident that the four independent sub-samples (quartiles) differ considerably from one another. However, the effect-sizes of all the sub-samples and the overall sample, distribute heterogeneously, as indicated by the χ^2 test of heterogeneity. The enormous heterogeneity also provides very good evidence for publication bias, which is by all means the most likely factor influencing effect-size distribution.

DISCUSSION

The publication of *ESP-60* marks the beginning of a new era. It stimulated laboratory research all over the Western world and helped parapsychology to gain general recognition. However, in comparison with current ESP research, the enormous *z*-scores stand out as extraordinary. In recent experiments such extraordinary *z*-scores have disappeared. For example, in a fundamental meta-analysis of Autoganzfeld studies, one of today's standard ESP experiments, the largest *z*-score among the 11 studies included is 3.04 (Bem & Honorton, 1994). Whether this indicates a fundamental ontological or phenomenological difference between early and recent experiments or whether this is an improvement of methodological quality has been the subject of much debate in parapsychology. However, the main argument in this paper is that two factors coincide and explain why the results of the experiments reported in *ESP-60* differ considerably from results achieved today. During the last 60 years dramatic changes have taken place (1) in the understanding of experimentation and (2) in the publication process in general.

ESP experiments developed from case studies at the end of the 19th century. Case studies at the time focused (as most case studies do today) on subjects with alleged extraordinary abilities (e.g. Gurney, Myers & Podmore, 1886; Crookes, 1889; James, 1896). Therefore, it is not surprising that almost all early ESP experiments and most experiments conducted at the beginning of the 20th century, used subjects with alleged extraordinary abilities, so called "gifted subjects". A fundamental principle of case studies is to examine phenomena in their natural setting or at least in a naturalistic setting. Because of this, systematic and tight controls are difficult to assure. However, this probably does not explain the statistically highly significant results of the early experiments. The reason rather is a second fundamental principle of case studies, i.e. no case will be filed (reported) without a (highly significant) phenomenon. Not reporting "insignificant cases" is not sloppiness or poor methodology, but part of the method. Case studies are not thought to be representative just as the experimenters who conducted the first experiments on ESP were by no means committed to a superior statistical perspective. The early experimenters were interested in demonstrating the phenomenon by reporting extraordinary data - they were not interested in meta-analysis. A purely statistical perspective on ESP might not even have convinced them about the phenomenon they were studying.

The advent of a more and more systematic and rigid experimental approach to ESP under Rhine's leadership in the late 1920s did not change the general attitude to the value of insignificant cases. Rhine strongly believed "that little can be learned from a report of an experiment that failed to find psi" (Broughton, 1987, p. 27). Although the Parapsychological Association has clearly rejected the policy of suppressing non-significant results since 1975 (Broughton, 1987; Honorton, 1985), it is still highly unlikely that experiments are published independently of their *p*-value. However, with the advent of meta-analysis it has become evident that conclusions based on unbiased samples are far more persuasive than conclusions based on biased samples, which might be challenged.

Publication bias is not a problem of ESP research in particular. Non-significant results are generally under-represented in published studies. There is no question that not every study is published and that journals are to different degrees filled with selective samples of statistically significant studies (e.g., Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995; Rosenthal, 1979). Especially when taking into account that the majority of studies published are underpowered, it is surprising that 95% of the respective articles in different psychological journals and more than 85% of the respective articles in different medical journals report statistically significant results (Sterling, Rosenbaum, & Weinkam, 1995). Editors, authors and reviewers are involved in the selection process.

However, the authors of *ESP-60* have been very prudent in addressing publication bias. They were probably the first to do so. They address the question by asking whether the "published [reports] give a complete or balanced view of the research accomplished" (p. 74). They were confident that publication bias is not a serious unilateral problem and argue that there might not only be a tendency not to publish data because it might for instance "not be worth anything" (p. 74), but that on the other hand there might also be a tendency not to publish data because "on the question of ESP, investigators obtaining positive results might not care to risk their reputations by publishing these results" (p. 75). To Pratt's et al. knowledge, thirteen studies with "favorable" results and seven with "chance results" were published suggesting "a tendency to suppress confirmatory results and to hasten to publish those which fail to confirm" (p. 76). However, this approach takes only into account studies known to the authors. The number of studies which are unknown to authors might be considerable, especially because those early experiments were relatively easy to conduct and not very cost and time intensive.

Pratt et al. have introduced a fail-safe approach to examining the impact of publication bias, which is very similar to Rosenthal's fail-safe approach (e.g. 1979). For their biggest sub sample of trials reported ($p = 1/5$, $n = 2{,}758{,}354$) they for instance calculated that it would require another 2,800,000,000 chance trials to reduce the z-score observed to 2.5 (p. 119). The fail-safe numbers for other sub samples such as for instance. reports published between 1934 and 1939 excluding sensory cues are equally high. However, the assumption that the unpublished trials are on average chance trials is questionable, especially when taking into account that the early experimental approach evolved from case studies which principally aimed at demonstrating ESP beyond statistical reasoning. The average *z*-score of unpublished studies might therefore be far from chance.

Up to this point the discussion focused on statistical significance, the main outcome parameter of the time. The analyses presented here were based on corresponding effect-size estimates considering the data from a different perspective. From this perspective publication bias seems to be the fundamental problem of the study sample. The effect-size distribution as shown in the funnel plot (see Figure 3) deviates strongly from the ideal distribution (theory). Effect-size estimates should scatter around the overall mean (see dotted line on Figure 3). They should scatter more widely at the bottom of the graph, with the spread narrowing among larger studies, because the precision in estimating effect-sizes increases with sample-size. To illustrate the deviation from the ideal funnel shape, a simulation based on the mean effect-size of the sample and the *ESP-60* sample-sizes has been carried out. Figure 4 shows how the funnel plot would ideally look like.

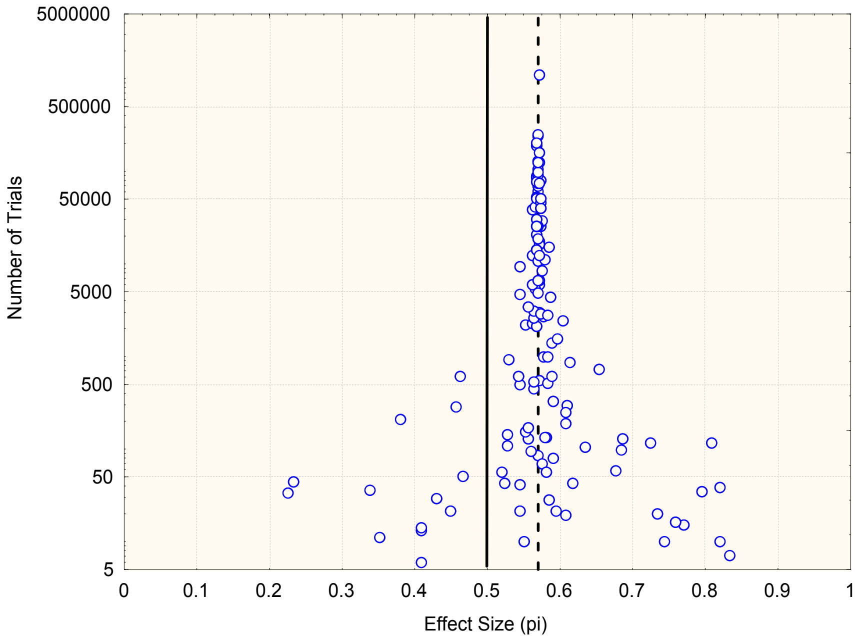


Fig. 4: Simulated – ideal – funnel plot of 139 studies with the mean effect-size as found in *ESP-60*.

From the difference between the ideal and the actual funnel plot it is evident that the data as collected by Pratt and his colleagues cannot be taken at face value. However, whether the asymmetry is actually a result of publication bias cannot be derived from the funnel plot. Several other biases, e.g. language bias, gray literature bias, multiple publication bias and database bias (for reviews of publication-related biases see e.g., Egger, Dickersin & Smith, 2001; Song, Eastwood, Gilbody, Duley & Sutton, 2000) can affect the effect-size distribution, as well as the thoroughness of the literature search (e.g. Song, Khan, Dinnes & Sutton, 2002). Skewed funnel plots can also be the result of effect-sizes varying in respect to a secondary variable like subject-experimenter relationship. The study sample might comprise of two or more independent sub-samples, which might lead to an unsymmetrical distribution of effect-sizes. Methodological quality can be related to effect-size and sample-size. Funnel plot asymmetry can also be the result of pure chance. However, even though some of the factors might play a role here, their impact on the effect-size distribution is likely to be negligible in comparison with publication bias, which seems evident here.

Whenever the funnel plot is asymmetrical it is likely that the asymmetry is due to the smaller studies in the sample, an effect which is called “small-study effect” (Sterne, Gavaghan & Egger, 2000). Actually, the impact of the smaller studies on the overall mean effect-size is immense. The overall mean effect-size, which normally converges in the bigger studies, does not converge due to the impact of the smaller studies in the sample. Two factors are generally considered to account for the small-study effect. Firstly, simulations demonstrate that publication bias (for statistical reasons) rather affects smaller than large studies (Hedges, 1984; Begg, 1994). Secondly, because smaller studies generally require fewer resources, the pressure to publish smaller studies independently of their results is much lower than for cost and time intensive larger studies (e.g. Begg & Berlin, 1988; Felson, 1992; Wilson & Henry, 1992; Copas & Shi, 2000).

To sum up, the discussions of the data, reasoning, and the analyses on the level of effect-size as well as on *z*-score suggest publication bias to be a fundamental problem of *ESP-60*. Publication bias clearly is a very plausible reason to account for the difference in effect-size between the smaller earlier studies and the subsequent larger studies. The smaller earlier studies were carried out to demonstrate dramatic effects (effect-sizes). At that time, the experimenters were not so much concerned with statistics, which with increasing sample size, later became the main concern of the experimenter. However, the fundamental conviction, i.e. that one cannot learn much from a statistically non-significant study, has not changed yet.

Conclusion

The switch in perspective in the data of the first comprehensive meta-analysis in the history of the empirical sciences revealed a striking funnel plot asymmetry and an extreme small-study effect. However, both approaches are purely descriptive and the question is whether publication bias alone explains this finding or whether other factors must also be considered. This question cannot be answered definitely. However, the publication of studies used to be based on a different perspective of the research process as a whole. Meta-analysis was not yet a common procedure and scientific proof at that time was obtained only by statistical significance, non-significant studies were generally put down to failure on the part of the experimenter or were attributed to other causes.

Pratt and his colleagues addressed publication bias, but they clearly underestimated the problem. In order to draw an overall conclusion, i.e. in order to carry out meta-analyses, a publication climate promoting the publication of non-significant studies must predominate or must be established. However, a different publication climate is not the only solution to the problem, especially because ESP experiments in general are not as cost and time intensive as for instance multicenter trials are in medicine, which nevertheless struggle with publication bias. Publication bias clearly is a fundamental problem of the empirical sciences.

Pratt and his colleagues have carried out and presented all the analyses with utmost care. Their meta-analysis was unmatched in its comprehensiveness until the last two decades of the 20th century, and although publication bias certainly challenges their primary conclusions, the empirical evidence they collected to support ESP is not automatically invalid just because it is challenged. To examine the alleged effect of this early period more closely, one would have to carry out a new met-analysis from the beginning, with a renewed literature search and a detailed coding of studies to examine the impact of potential moderator variables. However, modeling publication bias would be the major task of the meta-analysts. Without taking into account publication bias the enormous effect-size variability is probably not explicable.

For future ESP research I suggest registering studies and accepting only registered studies for publication. Publication bias cannot be prevented through well-meant statements. Such a policy would clearly minimize the impact of publication bias in future meta-analysis. Registers would not only help to provide a balanced perspective of ESP research altogether, but also and more importantly, underpin the quality of the empirical approach to ESP altogether.

ACKNOWLEDGMENTS

This research was sponsored by the Samueli Institute for Information Biology. I thank Ketevan Bachia and Anna Hack for retrieving the data and for doubling-checking them for accuracy. I also thank Andreas Sommer and two unknown referees for their valuable suggestions and comments. A table showing the effect-sizes and study characteristics for each study included in the meta-analysis will be available from Holger Bösch as soon as the article is accepted for publication in a peer reviewed journal.

REFERENCES

Alvarado, C. S. (1986). Letter. *Journal of* Parapsychology, *50*, 405-406.

Begg, C. B. (1994). Publication bias. In L. V. Hedges & H. Cooper (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Foundation.

Begg, C. B., & Berlin, J. A. (1988). A problem in interpreting medical data. *Journal of the Royal Statistical Society, Series A, 151,* 419-463.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088-1101.

Beloff, J. (1977). Historical overview. In B. B. Wolman (Eds.), *Handbook of Parapsychology* (pp. 3-24). New Nork: Van Nostrand Reinhold.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115 (1),* 4-18.

Bösch, H., & Walach, H. (2004a). The decline-effect, Part 1: Conceptual problems and explanatory approaches. *Manuscript submitted for publication,*

Bösch, H., & Walach, H. (2004b). The decline-effect, Part 2: Meta-analytic investigations. *Manuscript submitted for publication,*

Broughton, R. S. (1987). Publication policy and the Journal of Parapsychology. *Journal of Parapsychology, 51,* 21-32.

Broughton, R. S. (1991). *Parapsychology. The controversial science*. New York: Ballantine Books.

Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In B. B. Wolman (Eds.), *Handbook of Parapsychology* New York: Van Nostrand Reinhold.

Carington, W. W. (1935). Preliminary experiments in precognitive guessing. *Journal for Society for Psychical Research, 29,* 86-104.

Copas, J. B., & Shi, J. Q. (2000). Reanalysis of epidemiological evidence on lung cancer and passive smoking. *British Medical Journal, 320,* 417-418.

Crookes, W. (1889). Notes of séances with D. D. Home. *Proceedings of the Society for Psychical Research, 6,* 98-127.

Egger, M., Dickersin, K., & Smith, G. D. (2001). Problems and limitations in conducting systematic reviews. In M. Egger, G. D. Smith, & D. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (pp. 43-68). London: BMJ Books.

Egger, M., Smith, G. D., & O´Rourke, K. (2001). Rationale, potentials, and promise of systematic reviews. In M. Egger, G. D. Smith, & D. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (pp. 3-19). London: BMJ Books.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

Felson, D. T. (1992). Bias in meta-analytic research. *Journal of Clinical Epidemiology, 45,* 885-892.

Gurney, E. (1888). Recent experiments in hypnotism. *Proceedings of the Society for Psychical Research, 5,* 3-17.

Gurney, E., Myers, F. W., & Podmore, F. (1886). *Phantasms of the living*. London: Trübner & Co.

Hedges, L. V. (1984). Estimation of effect size under non-random sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics, 9,* 61-85.

Hedges, L. V. (1994). Fixed effect models. In L. V. Hedges & H. Cooper (Eds.), *The handbook of research synthesis* (pp. 285-299). New York: Russell Sage Foundation.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49,* 51-91.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science, 3,* 109-117.

James, W. (1896). Psychical research. *Psychological Review, 3,* 649-652.

Keil, J. (1987). Gaither Pratt, 1910 - 1979. In J. Keil (Eds.), *Gaither Pratt: A life for parapsychology* (pp. 1-59). Jefferson, NC: McFarland.

Lombroso, C. (1890). Thought-transference in Italy. *Journal of the Society for Psychical Research, 4,* 303-304.

Nash, C. B. (1989). Intra-experiment and intra-subject scoring declines in 'Extra sensory Perception After Sixty Years'. *Journal of the Society for Psychical Research, 55,* 412-416.

Parapsychological Association. (18-2-2004). Who was J. B. Rhine? URL: *http://www.parapsych.org/members/jb_rhine.html* [18-2-2004].

Pratt, J. G., Rhine, J. B., Smith, B. M., Stuart, C. E., & Greenwood, J. A. (1940). *Extra-sensory perception after sixty years: A critical appraisal of the research in extra-sensory perception.* New York: Henry Holt and Company.

Radin, D. I. (1997). *The conscious universe*. San Francisco: Harper Edge.

Rao, K. R. (2001). Introduction: Reality, replicability, and lawfulness of psi. In K. R. Rao (Eds.), *Basic research in parapsychology* (pp. 3-37). Jefferson, NC: McFarland.

Rhine, J. B. (1934). *Extrasensory perception* . Boston: Boston Society for Psychic Research.

Rhine, J. B. (1942). Evidence of precognition in the covariation of salience ratios. *Journal of Parapsychology, 6,* 111-143.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Rosenthal, R., & DiMatteo, R. M. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59-82.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106,* 332-337.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration, 14,* 91-106.

Schlitz, M., & Gruber, E. (2001). Transcontinental remote viewing. In K. R. Rao (Eds.), *Basic research in parapsychology* Jefferson, NC: McFarland.

Shield Mrs, M. E. (1887). Experiments in thought transference. *Journal for Society for Psychical Research, 3,* 179-182.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., & Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessment, 4,* 1-115.

Song, F., Khan, K. S., Dinnes, J., & Sutton, A. J. (2002). Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology, 31,* 88-95.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association, 54,* 34.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician, 49,* 112.

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53,* 1119-1129.

Titchener, E. (1898). The feeling of being stared at. *Science, 8,* 895-897.

White, R. A. (1991). The psiline database system. *Exceptional Human Experience, 9,* 163-167.

Wilson, A., & Henry, D. A. (1992). Meta-analysis. Part 2: Assessing the quality of published meta-analyses. *The Medical Journal of Australia, 156,* 173-80, 184.

Address for correspondence: Holger Bösch,
Freiburg University Hospital, Department of
Environmental Medicine and Hospital Epidemiology,
Hugstetter Str. 55, D-79106 Freiburg, Germany.
E-mail: hboesch@ukl.uni-freiburg.de

EXPLORING THE RELIABILITY OF THE "PRESENTIMENT" EFFECT

Richard S. Broughton
Intuition Laboratories, Inc., Durham

ABSTRACT

This project extends a series of experiments begun by Radin that appears to demonstrate an autonomic nervous system (ANS) response to future emotionally arousing experiences. This experiment used skin conductance as the ANS measure and was designed to examine test-retest reliability in the "presentiment" effect, also known as "pre-stimulus response" (PSR). The experiment also looked for individual differences in PSR responding using the Myers-Briggs Type Indicator and the NEO-FFI personality inventories.

Volunteer participants completed two sessions of viewing 40 calm and arousing images from the International Affective Picture System (IAPS) while SC was monitored. The PSR window was defined as 3 seconds before stimulus onset, and the measure of PSR was the area between the mean relative SC levels for the two classes of stimuli within the prestimulus window. Significance was determined using a Monte Carlo method. A novel analysis technique of de-trending the SC data immediately before the PSR analysis window was used to counter the possible effect of expectancy and other artifacts. Eighty planned participants were tested. Sixteen were eliminated from further analysis because of lack of any SCRs, leaving 64 participants in the final data pool.

The experiment did not produce overall evidence of a presentiment effect, nor did it demonstrate test-retest reliability. Since this experiment deviated from prior methods of analysis the data were examined without de-trending and automatic artifact removal, but there was still no evidence of a presentiment effect. Of three personality factors previously associated with better ESP scoring, MBTI Extraversion, MBTI Intuition, and NEO Openness, the last two were positively and significantly correlated with individual mean PSR. The data were explicitly tested for the presence of an expectancy artifact by correlating the PSR prior to arousing stimuli with the time since the last arousing stimulus. The correlations for both the de-trended data and the original method of analysis were negligible and non-significant, providing no evidence that an expectancy artifact was present.

INTRODUCTION

Finding evidence of physiological indicators of the operation of extrasensory perception has long been a holy grail of parapsychology (Dean, 1962, 1966; Hartwell, 1978; Levin & Kennedy, 1975; Tart, 1963). Recently, Radin introduced a new method of using physiology to detect possible precognition of future emotionally shocking experiences. Termed "presentiment," Radin's technique makes use of the responses comprising the orienting response (or "flight or fight" response), most commonly the skin conductance response (SCR). In a typical experiment, a participant's physiology is monitored while he or she views a random series of calm or arousing images. In this design, frequently used in emotion research, a participant usually produces a pronounced response, e.g., a rise in skin conductance, about two seconds after a shocking picture appears. Radin, however, looked at the period several seconds *before* the picture was shown and he found that for some participants their skin conductance began rising before a shocking picture was shown, but not before a calm picture. It appeared that the participant's nervous system was anticipating—preparing for, perhaps—a future emotional shock or threat.

Radin reported this finding in 1997 (Radin, 1997) and since then a series of increasingly sophisticated experiments has yielded reasonably consistent results (for a summary see Radin, 2003). Other researchers have replicated the findings (Bierman & Radin, 1997; Bierman & Scholte, 2002). Using a similar design but with a startle response (loud noise) rather than an emotional image, May and Spottiswoode have obtained strikingly successful results (May & Spottiswoode, 2003; Spottiswoode & May, 2003), which they called "pre-sentiment response" (PSR). Very recently, however, the investigators have raised questions about the interpretation of their findings as representing a physiological response by the participant (May, 2004).

As with any new experimental technique in parapsychology, questions have been raised about possible artifacts. At the outset it was recognized that there could be an expectancy artifact, a "gamblers fallacy" of the autonomic nervious system, in which the participant's skin conductance rises more or less monotonically between trials with shocking stimuli. The longer the participant waits, the greater the expectation of a shocking picture and the higher the skin conductance level. Initial considerations indicated this was not a problem (Bierman, 1999; Radin, 1997). Later Dalkvist, Westerlund and Bierman (2002) demonstrated that the expectancy bias remains a problem through elaborate simulations of PSR studies in which the data of individual test sessions were combined in different ways. Dalkvist et al. observed, however, that when data are averaged for the arousing and calm conditions across all participants the bias is not a problem in practice, though it does not disappear completely. Wackermann (2002) has provided a mathematical treatment of the bias. Radin has reexamined some of his past data in the light of these analyses and found no evidence of an expectancy artifact in those experiments (Radin, 2003). For their recent experiments Spottiswoode and May developed an analysis method that controls for the possibility of expectancy artifact (Spottiswoode & May, 2003) as well as a possible position artifact.

Elsewhere I have drawn attention to the range of spontaneous cases where the response by the recipient of the anomalous information is an emotional or feeling response (Broughton, 2002). Stevenson (1970) has studied a number of these cases in detail, termed intuitive cases, and he notes that they often involve appropriate emotional and behavioural responses but little or no imagery of other cognitive content. The most frequent source of spontaneous ESP is in dreams, and these often involve threats to relationships with loved ones (Ullman, Krippner, & Vaughan, 1989), as do many of the spontaneous cases in general (Schouten, 1982). Recent PET studies of dreaming have shown that the REM state of dreaming seems to activate a tight circuit involving the limbic system and visual association areas, but not the primary visual area of the brain, while the prefrontal cortices show decreased activity (Braun et al., 1998).

Coupled with suggestions from experimental research there is reason to suspect that the emotional system of the brain and body play a role in the transduction of anomalous information into appropriate and adaptive responses by the recipient (Broughton, 2002), possibly in an evolutionarily feasible manner similar to Damasio's somatic markers (Damasio, 1994, 1996).

It is tempting to speculate that the presentiment response possibly reflects a fundamental process by which anomalous information becomes an intuition or an intuitive decision that proves to be useful. As a first step toward linking presentiment with intuitive decision-making it is necessary to explore the reliability of presentiment as a human response. If presentiment is reliable—a type of "intuition response"—then it would also be reasonable to expect individual differences in the intuition response. Thus, the primary objective of this project was to establish the statistical reliability of the presentiment response on a sufficiently large population using the skin conductance response obtained by accepted and widely practiced methods. A secondary objective was to demonstrate measurable individual differences in the presentiment response by examining whether it correlates with selected psychological instruments.

METHODS

Overview

The general design of this experiment involved having volunteer participants participate in two sessions of a presentiment experiment, using a methodology similar to that of Radin (Radin, 1997, 2000) and Bierman (Bierman & Radin, 1997). Tests were administered typically at one-week intervals, though that time varied according to individual requirements. Participants also completed the Myers-Briggs Type Indicator (MBTI) and the NEO-FFI personality assessments, prior to the presentiment testing in most cases.

Participants

Eighty-three participants were recruited and tested. Three were discarded before analysis due to equipment failures early in the series, leaving the target number of 80 participants. Upon initial inspection of the skin conductance data, 16 participants were found to have produced no responses in one or both sessions, i.e., no apparent skin conductance responses (SCRs) in a full session. These participants were also discarded, leaving 64 qualifying participants. The qualifying participants consisted of 25 males and 39 females, ranging in age from 19 to 81 and with a mean age of 44.7 (SD = 14.8). Participants were recruited primarily via word-of-mouth or appeals by the experimenter at various meetings or groups of interested persons. Participants were not paid, but were given a formal report of their MBTI results for personal use. All participants executed an informed consent form prior to participation.

Apparatus

Skin conductance was measured using a PSYLAB model SC5-SA monitor and pre-amplifier system manufactured by Contact Precision Instruments of London, U.K. and Cambridge, MA[1]. The SC5 is a 24 bit digitizing monitor with 0.1 µSiemens absolute accuracy, better than 0.001 µSiemens relative accuracy, and a range of 100 µSiemens. It measures directly in units of conductance, using DC coupling with constant voltage excitation.

The SC5 unit samples at 40 Hz and outputs its data in real time via a standard RS232 serial connection to a computer.

Ag/AgCl electrodes of 8 mm diameter (MED Associates model TDE-022-48) were filled with an isotonic electrode paste (MED Associates type TDE 246 Skin Conductance Electrode Paste) equivalent to Lykken and Venables' (Lykken & Venables, 1971) "Unibase".

Emotional pictures were drawn from the International Affective Picture System (IAPS) (Lang, Bradley, & Cuthbert, 1995) and presented on the screen of a laptop computer as described below. To enhance the contrast between the calm and arousing images, only the top 220 and bottom 220 pictures on the arousal scale were used. There are separate scales for males and females in the IAPS system.

A laptop computer (Dell Inspiron 3800 with Windows 98 SE operating system) was used both to present the images to the participant and to collect the skin conductance data. The controlling program was written in Visual Basic 6 by Edwin May of the Laboratories for Fundamental Research and independently tested and modified by the author. Picture selection was randomized using a hardware-based "true" random number generator (model rp1)[2] attached to the parallel port. If the hardware RNG was not available the program defaulted to a fail-safe mode that employed a high-quality pseudo-RNG (Marsaglia & Zaman, 1987). Both RNG systems passed Marsaglia's DIEHARD randomness test and both were tested to confirm adequate randomization of the IAPS image set as used in this experiment. The procedure for determining pictures involved two steps. In the first step both a calm and an arousing picture were selected from the available pool and loaded into memory. Immediately prior to the stimulus presentation another random decision determined which of the two pictures would appear. Thus the decision as to whether to show a calm or arousing picture was made only about a microsecond before the picture appeared on the screen and there was no computer disk movement or other potential source of noise associated with the random decision. The very short time between the random decision and the appearance of the picture eliminates the possibility of any cueing artifacts arising from the equipment.

Picture stimuli were presented to the participant for 3 seconds each with an inter-stimulus interval that varied randomly between 21 and 26 seconds. Images appeared approximately 13 x 17 cm in the center of a black screen.

---

[1] Contact Precision Instruments Inc., P.O. Box 2603, London N1, UK. http://www.psylab.com/ .

[2] Rolf Freitag, University of Ulm, Dept. of Semiconductor Physics, Albert-Einstein-Allee 45, 89069 Ulm, Germany. http://hlhp1.physik.uni-ulm.de/~freitag/spinoffs.html.

Skin conductance data were collected continuously during the experimental period via standard Windows API routines. Input buffers were transferred to program data arrays at critical points in the program, for example, immediately prior to picture selection, for accurate synchronization of SC data with the picture events.

Personality assessment

The Myers-Briggs Type Indicator (MBTI) Form G was used with standard response sheets that were scored by a program written by the experimenter. The scoring program checked for data entry integrity and stored results in a database. The MBTI was selected because of a history of demonstrated relationships between MBTI factors and ESP performance.

The NEO-Five Factor Inventory (NEO-FFI) Form S was also used. These were hand-scored by the experimenter. The NEO-FFI has shown some correlations between ESP performance and at least the "Openness to experience" scale and it has also been used in studies of emotional responses to the IAPS picture set.

Both personality tests were normally given to participants and completed prior to the IAPS test sessions although there were a limited number of exceptions to this.

Testing environment

In order to maximize the number of participants, testing was offered with considerable flexibility in location and time. The equipment was portable, so, in addition to the facilities of Intuition Laboratories, participants were tested in their homes or those of friends, and sometimes at their place of work. In some instances, a participant would organize a small group of friends to chat with the experimenter and take part in the experiment. In all cases, the actual test took place in a quiet room (away from any other people, if there were any) with comfortable viewing arrangements. In all cases, participants took both the first and second tests in the same environment. Under these circumstances it was not possible to control closely temperature and humidity, but in all cases these were within the normal "comfort zone" for American households (20° to 25° C).

Testing procedure

After preliminary interactions, the experimenter would lightly wipe the first and second fingers of the participant's non-dominant hand with an ethanol wipe. The prepared electrodes would be fixed to the medial phalanx of these fingers with paper medical tape in a cross fashion to minimize constriction. A minimum of 10 minutes was allowed to elapse during which the experimenter explained the experiment and generally discussed its importance. The participant was seated in a comfortable chair at a convenient viewing distance from the computer screen. The experimenter then loaded the parameter file into the program and entered various session details. Next the program displayed a graph of the skin conductance and the experimenter provided a brief explanation and a demonstration by asking the participant to take a sharp, deep breath (which typically produced a noticeable deflection in the trace). The experimenter then reviewed the sequence of events that would follow and allowed the participant to experience a demonstration trial (always a calm image that was not used in the actual test). When the participant had no further questions, the experimenter reminded the participant to simply sit quietly and watch the pictures, breathing normally and avoiding any large bodily movements if possible. Then the experimenter started the program and left the testing room. At that point the program began collecting SC data and displaying the 40 randomly selected IAPS pictures. The testing time lasted about 18 minutes and on completion a message was displayed thanking and instructing the participant that the experiment was completed and to call the experimenter. No feedback of results was provided at this time, but the experimenter invited the participant to express his or her reactions to the material if they wished and answered any additional questions.

The second test session was held about a week later, although in some cases the interval was longer and, on rare occasion, shorter, when this was necessary to insure that the participant could complete both sessions. The time interval was not thought critical at this stage of the research. For the second session the explanation of the experiment was in the form of reminding the participant what would take place and the demo trial was not used. Each test session used a fresh random sequence of images, so participants saw a different series of pictures on the second visit (apart from any that might have randomly come up in the first session).

After the second test session was completed the experimenter provided the participant with his or her MBTI and NEO-FFI reports with a short explanation of the results.

Hypotheses and planned tests

The two principal hypotheses were (1) that the combined data of all sessions would demonstrate significant evidence of a pre-stimulus effect using the epoch analysis method (difference in area between the arousing SCR level and the calm SCR level in the prestimulus window), and (2) that there would be a positive correlation between the prestimulus responses of the two test sessions for the participants. Hypothesis 1 would demonstrate a replication of previous work and hypothesis 2 is expected on the basis of the nature of human abilities (although replicability in psychological tests is not as reliable as some suppose (see, for example, Kindt, Bierman, & Brosschot, 1996).)

It was hoped that one or more of the various personality factors derived from the two tests would show a relationship with participants' average prestimulus response. In previous standard ESP testing, MBTI factors E (extraversion) and N (intuition) have been shown to be associated with positive scoring (Broughton, Kanthamani, & Khilji, 1989; Honorton et al., 1990), and ganzfeld testing had shown persons with the MBTI profile of ENFP to be good scorers (Honorton, 1992). In the NEO-FFI, only the O (openness to experience) had been associated with good ESP scoring (Broughton & Alexander, 1996). It was, therefore, expected that these personality factors might show a positive relationship with PSR.

Data reduction and analysis

Data reduction and analysis was conducted in close collaboration with Edwin May and James Spottiswoode of the Laboratories for Fundamental Research who were engaged on a similar project using an audio startle stimulus. They introduced the term pre-stimulus response since their stimuli had no affective content, and this term will be used for this experiment.

The original method of analyzing presentiment experiments was to average all arousing trials for all participants and compare the resulting averaged response with the averaged response for the calm trials. Prior experiments defined the prestimulus period as 5 seconds before the start of the stimulus presentation. For a number of reasons, we decided not to follow exactly the original analysis method.

To deal with the possibility of expectation and position artifacts, discussed in the introduction, Spottiswoode and May (2003) developed a method of analysis in which any trend in the data (up, for example, in the case of expectancy, or down in the case of position artifacts) is removed through a curve-fitting technique immediately prior to the pre-stimulus window. With the trend removed, data are clamped to zero at the start of the prestimulus window and the prestimulus response is defined as the area under the curve defined by the relative skin conductance difference. A detailed explanation of this analysis method can be found in Spottiswoode and May (2003).

Pilot explorations with data from similar PSR experiments led to the specification of two additional thresholds for inclusion of PSR data. The pilot data revealed that the PSR "effect," which, if real, is very small, and is very sensitive to non-specific SCRs (NS.SCR) that can occur normally throughout the session, due to the participant moving slightly or just taking a deeper than normal breath. It is not possible to identify and remove these artifacts manually, so two criteria were developed for automatic exclusion. Any changes in the slope of the SCR line within the prestimulus area that fell within the top 2.5% of positive or negative deflections (most extreme 5% of changes in slope) caused the trial to be discarded. Similarly, any shifts in skin conductance level that fall within the top 2.5% of positive and negative excursions (5% total) were excluded as likely artifacts. The cutoffs were derived by examining all slope changes and level changes in the qualifying trials, excluding the prestimulus region and 10 seconds following the presentation of the stimulus. For further details see Spottiswoode and May (2003).

The pilot explorations also revealed that the most efficient window in which to look for a presentiment effect using the curve-fitting approach was 3 seconds before the stimulus onset. This is different from the PSR window used by previous researchers who typically used a 5 second window.

Post-stimulus response amplitudes were computed in the standard way. Only SCR amplitude was used for examination of post stimulus responses.

Lability measures for each participant were calculated by counting the number of NS.SCRs outside the prestimulus window and 10 seconds post stimulus and dividing this by the number of seconds, yielding a NS.SCR/second value.

Spottiswoode and May (2003) have also developed an additional measure for examining prestimulus response. This is simply to count the number of NS.SCRs in the prestimulus window and compute a simple ratio of the average number of NS.SCRs for the arousing condition over the average number of NS.SCRs for the calm condition. The resulting ratio is tested against the expected ratio of 1.0. Spottiswoode and May argue that this may be a better method of analysis for PSR.

The planned hypotheses were to be tested as follows. Hypothesis 1, that there would be overall evidence of a PSR effect, would be tested by averaging the arousing trials and the calm trials for all participants using the curve-fitting method in the pre-stimulus window of −3 to 0 seconds and computing the difference in area under the curves. This is called the prestimulus area difference (PAD) and the analysis method is also known as an epoch analysis. The significance of that difference would be determined by a 20,000 pass Monte Carlo simulation that randomly re-assigned arousing and calm labels to responses. Hypothesis 2, that there would be a positive correlation between participant performance across the two test sessions would be tested using a Spearman correlation on the z-score based on the PAD for each test session.

RESULTS

Hypothesis 1 was not confirmed. In fact, there was no evidence of a PSR effect using the planned measure. The Monte Carlo analysis of the overall PAD from the 128 sessions (2 from each of the 64 qualifying participants) yielded a $z = 0.26$ (ns). Figure 1 presents the results.

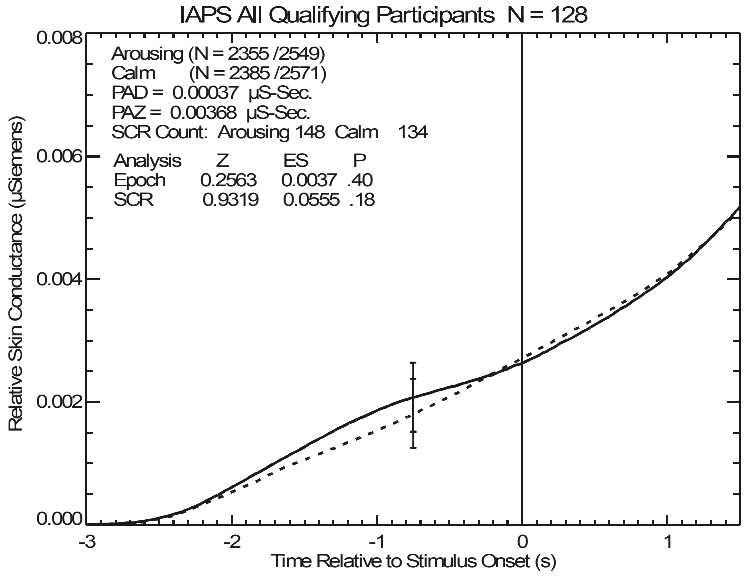


Figure 1: Epoch analysis of 128 sessions. Dashed line is for calm stimuli. N's for each condition represent the number of trials after filtering out of the total number of trials.

Spottiswoode and May's alternative method of evaluation based on the NS.SCRs in the prestimulus window yields a slightly better, but still completely non-significant $z = 0.93$.

Hypothesis 2, that there would be a correlation between participants' test sessions, as measured by their individual PAD scores, was not confirmed either, $r = -.12$ ($df = 62$, $p = .35$).

To examine the relationship between individual PSR results and the personality tests, a mean z was computed based upon the z-scores associated with the two PAD results for each participant. The continuous scores from the four factors of the MBTI test and the T-scores from five factors of the NEO-FFI were used. Of the nine items, the MBTI SN (sensing-intuition) scale showed a positive and significant correlation with PSR, $r = .29$ ($df = 62$, $p = .021$), as did the NEO Openness scale, $r = .32$ ($df = 61$, $p = .012$). Table 1 shows all correlations.

Table 1 Personality factor correlations with pre-stimulus response

Factor	Pearson r (p)
MBTI Extraversion-Introversion	–.13 (.29)
MBTI Sensing-Intuition	.29 (.02)
MBTI Thinking-Feeling	.05 (.70)
MBTI Judging-Perceiving	.08 (.52)
NEO Neuroticism	–.14 (.26)
NEO Extraversion	.07 (.57)
NEO Openness	.32 (.01)
NEO Agreeableness	.04 (.74)
NEO Conscientiousness	.04 (.73)

Notes: MBTI tests have 62 df, NEO have 61 df (one person did not complete). MBTI E-I scale puts introversion high, thus a negative sign indicates a positive correlation with extraversion. Probabilities are not corrected for multiple analyses.

In light of the comments by Kindt et al. (Kindt et al., 1996) that test-retest reliability in psychology is frequently assumed rather than demonstrated, it seemed appropriate to test the reliability of the *normal* components of the SC response to calm and arousing IAPS pictures. To assess normal reliability I computed the mean relative amplitude of the SCRs for calm and arousing stimuli and the difference between the mean amplitudes for each test session. These measures from the two sessions of each participant were tested in Pearson correlations ($df = 62$ in all cases): Calm amplitude, $r = .47$ ($p = .00009$); Arousing amplitude, $r = .44$ ($p = .0002$); Amplitude difference, $r = .32$ ($p = .009$).

Decomposing the PSR

In a similar experiment using loud audio stimuli for the arousing condition, May and Spottiswoode (May & Spottiswoode, 2003; Spottiswoode & May, 2003) have demonstrated that the principal component of the anomalous PSR is one or more NS.SCRs that occur in the pre-stimulus region of each trial, not the gradual rise in SCL that seems to be the effect in prior research. The apparent rise in SCL may simply be the consequence of averaging many NS.SCRs that occur more frequently before arousing than calm stimuli. To examine the effect of NS.SCRs in the prestimulus region, I compared the basic epoch analysis for those sessions in which there was at least one NS.SCR in the prestimulus region of either a calm or an arousing stimulus with the same analysis done on sessions with no NS.SCRs in the prestimulus regions. The results are shown in Figure 2.

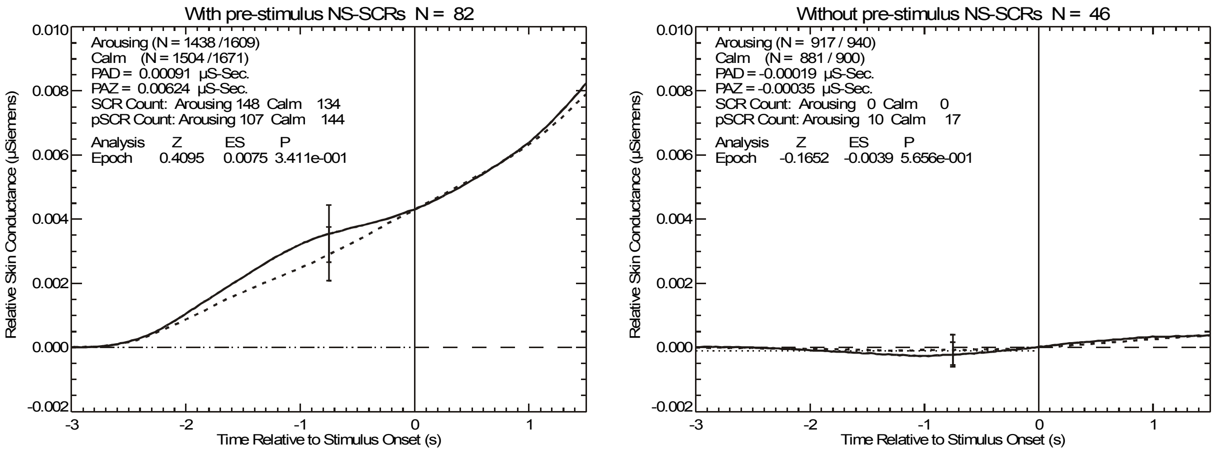


Figure 2: Epoch analysis using trials with NS.SCRs in prestimulus region (left), without NS.SCRs (right). Dotted line represents calm stimuli.

Inspection of the two plots reveals a striking difference. When there are no NS.SCRs in the prestimulus region, there is no increase in SCL prior to either calm or arousing trials.

EXAMINING NORMAL EXPLANATIONS

The overall results of the experiment did not provide a significant confirmation of the hypotheses, yet it is still prudent to examine potential normal explanations for any results obtained.

Randomness of stimuli

Although the true RNG passed exhaustive acceptance tests, it is also necessary to examine its performance in the actual runs used in the experiment. A Wald-Wolfowitz Runs test was applied to the calm-arousing sequences for each of the 128 sessions used. Of the 128 *z*-score results, four were significant, three exceeding the .05 level of significance and 1 falling at the .01 level, which is within normal expectation, indicating that the stimuli conditions were allocated randomly.

Normal expectancy (gambler's fallacy)

Although the experiment did not provide evidence for PSR there was no reason to suppose that it should be any less susceptible to the expectancy artifact. To test for evidence of an expectancy artifact I computed a prestimulus area from zero (PAZ) score representing the area between the SC line and the zero clamping line in the prestimulus region for each trial. For the arousing trials only, I computed a Spearman correlation between the PAZ for the arousing trials and the number of seconds since the last arousing trial (or the start of the experiment where necessary). If an expectancy artifact is present, it should manifest as a positive correlation between PAZ and the time between arousing stimuli.

For the data as used in this experiment, which was de-trended to remove any expectation bias and filtered to remove potentially artifactual NS.SCRs, the result was r_s = .016 (df = 2353, $p_{1\text{-}tail}$ = .21). To test the potential effect of an expectancy artifact on more typical data, the PAZ score was computed without de-trending or artifact filtering and used in the Spearman correlation. The result was r_s = .003 (df = 2547, $p_{1\text{-}tail}$ = .44). This represents a clear test of the expectancy effect hypothesis. Finally, May and Spottiswoode's alternative method of analysis using the ratio of NS.SCRs was tested by using the count of NS.SCRs before an arousing stimuli instead of the PAZ score. The result was r_s = .013 (df = 2547, $p_{1\text{-}tail}$ = .25). In all three tests there was no evidence of an expectancy artifact.

DISCUSSION

This experiment clearly failed to replicate previous experiments that have used a similar design. Since there were a number of differences in the way this experiment was analyzed when compared with previous work, an obvious question is whether the original method of analysis would have yielded better results. The original method used a 5 second PSR window and used simple ensemble averaging rather than the curve-fitted de-trending used here. The simple answer is no, as shown in Figure 3.

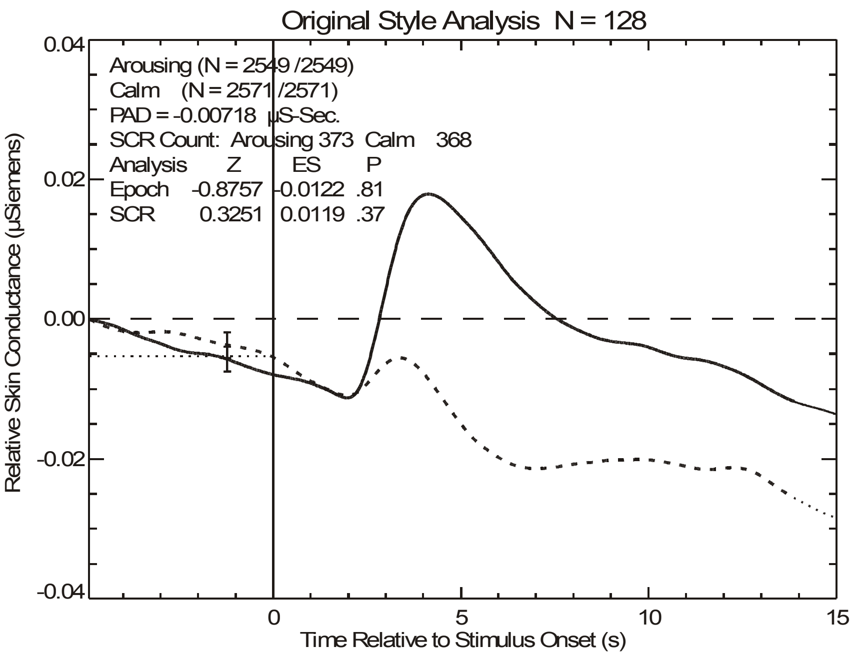


Figure 3: Results of this experiment if it had been analyzed using the "traditional" method. Dashed line represents calm stimuli.

That the experiment failed to demonstrate test-retest reliability is not surprising since there was no primary effect.

Several possibilities may be considered for why this experiment failed to replicate previous findings.

One possibility is that there was insufficient contrast between the calm and arousing pictures. In previous experiments there has not been uniformity in stimuli selection. Some previous experiments have used IAPS images mixed with similar images from unspecified sources, while others have explicitly replaced some of the IAPS pictures (erotic images, for example) with more extreme versions tailored for local conditions. In this experiment I deliberately wanted to stay with the well-tested standard set. It is clear from Figure 3 that the IAPS images had their intended effect on the normal post-stimulus response, so it would be difficult to argue that they were not appropriate for testing the prestimulus effect.

In the similar experiments using audio startle stimuli May has found strong circumstantial evidence that the striking results were due to an experimenter effect (May, 2004). In this scenario, the experimenters used their precondition to begin each test session at the optimal moment to obtain the best fit between the naturally occurring NS.SCRs and the random sequence of stimuli and timing that the computer would generate. Under this analysis the present experimenter simply failed to produce.

On the positive side, of the nine personality factors investigated in this experiment, three have previously been identified as being associated with better scoring in ESP experiments. All three were correlated in the expected direction and two of them, MBTI Intuition and NEO-FFI Openness, were correlated significantly with PSR. Even as one-tailed probabilities they would not survive a correction for multiple analyses, however this result is a promising indication that experiments with more robust evidence of PSR may reveal personality relationships consistent with previous ESP research.

The experiment confirmed what may prove to be an important observation by Spottiswoode and May (2003) that the PSR or presentiment effect may consist of an excess of NS.SCRs in the pre-stimulus region rather than a gradual rise in skin conductance level prior to arousing stimuli that has been presumed to be the basic effect in prior research. Future experiments should plan to examine this issue systematically.

One of the useful findings to emerge from this experiment is that it provides strong evidence that the expectation artifact may not be a serious problem in practice. The complete lack of even a suggestion of a correlation between arousing trial PSRs and the duration from one arousing to the next arousing stimuli indicates that the participants in this experiment were not displaying the anticipation strategies proposed by Dalkvist et al. (2002).

Although this experiment failed to confirm the hypotheses it set out to test, it has managed to apply some promising new tools for the continuing investigation of the prestimulus response or presentiment to emotionally charged materials. Future research should address the relationship between the NS.SCRs and the gradual SCL increase previously taken as the indicator of PSR, as well as the degree to which individual differences play a role in the results.

ACKNOWLEDGEMENTS

Intuition Laboratories, Inc. gratefully acknowledges the support of the Bial Foundation, Porto, Portugal, for this experiment under Bursary 52/00. I wish also to acknowledge and thank Dr. Ed May of the Laboratories for Fundamental Research, Palo Alto, CA, for supplying the data collection and data analysis programs, and for invaluable advice throughout this project. Thanks also to the many participants, especially those who helped with the recruiting.

References

Bierman, D. J. (1999). *Sequential patterns cannot explain anomalous anticipatory physiological behavior: A proof*. Retrieved 30 July, 2002, from the World Wide Web: *http://www.psy.uva.nl/resedu/pn/pubs/bierman/1999/proof.html*

Bierman, D. J., & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills, 84*, 689–690.

Bierman, D. J., & Scholte, H. S. (2002, August 5–8). *Anomalous anticipatory brain activation preceding exposure to emotional and neutral pictures.* Paper presented at the The Parapsychological Association, 45th Annual Convention, Paris.

Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwadry, F., Carson, R. E., Varga, M., Baldwin, P., Belenky, G., & Herscovitch, P. (1998). Dissociated Pattern of Activity in Visual Cortices and Their Projections During Human Rapid Eye Movement Sleep. *Science, 279*(5347), 91–95.

Broughton, R. S. (2002). Telepathy: Revisiting its Roots, *4º Simpósio da Fundação Bial: Aquém e Além do Cérebro: Relações Interpessoais Excepcionais.* (pp. 131–146). Porto: Fundação Bial.

Broughton, R. S., & Alexander, C. H. (1996). *Autoganzfeld II: An attempted replication of the PRL ganzfeld research.* Paper presented at the Parapsychological Association 39th Annual Convention, San Diego, CA.

Broughton, R. S., Kanthamani, H., & Khilji, A. (1989). *Assessing the PRL success model on an independent ganzfeld data base.* Paper presented at the Parapsychological Association 32nd Annual Convention, San Diego, California.

Dalkvist, J., Westerlund, J., & Bierman, D. J. (2002, August 5–8). *A computational expectation bias as revealed by simulations of presentiment experiments.* Paper presented at the The Parapsychological Association 45th Annual Convention, Paris.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G. P. Putnam's Sons.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London, B, 351*, 1413–1420.

Dean, E. D. (1962). The Plethysmograph as an indicator of ESP. *Journal of the Society for Psychical Research, 41*, 351–353.

Dean, E. D. (1966). Plethysmograph recordings as ESP responses. *International Journal of Neuropsychiatry, 2*, 439–446.

Hartwell, J. W. (1978). Contingent negative variation as an index of precognitive information. *European Journal of Parapsychology, 2*, 83–103.

Honorton, C. (1992). *The ganzfeld novice: Four predictors of initial ESP performance.* Paper presented at the Parapsychological Association 35th Annual Convention, Las Vegas, NC.

Honorton, C., Berger, R. E., Varvoglis, M. P., Quant, M., Derr, P., Schecter, E. I., & Ferrari, D. C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology, 54*(2), 99–139.

Kindt, M., Bierman, D., & Brosschot, J. F. (1996). Stroop versus Stroop: Comparison of a card format and a single-trial format of the standard color-word Stroop task and the emotional stroop task. *Personality and Individual Differences, 21*(5), 653–661.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1995). *International Affective Picture System (IAPS): Technical manual and affective ratings.* Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

Levin, J., & Kennedy, J. (1975). The relationship of slow cortical potentials to psi information in man. *Journal of Parapsychology, 39*, 25–26.

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology, 8*, 656–672.

Marsaglia, G., & Zaman, A. (1987). *Toward a universal random number generator* (FSU-SCRI-87-50). Gainesville, FL: Florida State University.

May, E. C. (2004). *Anticipatory skin conductance responses: A possible example of decision augmentation theory. (Submitted).* Paper presented at the Parapsychologocial Association 47th Annual Convention, Vienna Austria.

May, E. C., & Spottiswoode, S. J. P. (2003). *Skin conductance response to future audio startle stimuli.* Paper presented at the Parapsychological Association 46th Annual Convention, Vancouver, Canada.

Radin, D. I. (1997). Unconscious perception of future emotions: An experiment in presentiment. *Journal of Scientific Exploration, 11*(2), 163–180.

Radin, D. I. (2000, 7/19/00). *Evidence for an anomalous anticipatory effect in the autonomic nervous system.* Boundary Institute. Retrieved July 22, 2002, from the World Wide Web: *http://www.boundaryinstitute.org/articles/presentiment99.pdf*

Radin, D. I. (2003). *Electrodermal presentiments of future emotions.* Paper presented at the Parapsychological Association 46th Annual Convention, Vancouver, Canada.

Schouten, S. A. (1982). Analyzing spontaneous cases: A replication based on the Rhine collection. *European Journal of Parapsychology, 4*, 113–158.

Spottiswoode, S. J. P., & May, E. C. (2003). Skin conductance prestimulus response: analyses, artifacts and a pilot study. *Journal of Scientific Exploration, 17*(4), 617–641.

Stevenson, I. (1970). *Telepathic impressions: A review and report of thirty-five new cases.* Charlottesville, VA: University Press of Virginia.

Tart, C. T. (1963). Possible physiological correlates of psi cognition. *International Journal of Parapsychology, 5*, 375–386.

Ullman, M., Krippner, S., & Vaughan, A. (1989). *Dream telepathy: Experiments in nocturnal ESP.* Jefferson, N.C.: McFarland.

Wackermann, J. (2002, August 5–8). *On cumulative effects and averaging artifacts in randomized S-R experimental designs.* Paper presented at the The Parapsychological Association 45th Annual Convention, Paris.

Address for correspondence: Richard S. Broughton,
Intuition Laboratories, Inc., Durham,
North Carolina 27715, USA.
E-mail: richard@intuitionlabs.com

RULING OUT ITEMS IN THE JUDGING SET: A NOVEL PROTOCOL FOR DETECTING PSI IN THE GANZFELD

Jezz Fox and Christine Simmonds
Psychology, Liverpool Hope University College

ABSTRACT

This paper considers whether receivers may find it easier to identify items in the judging set that are not the target rather than taking the more traditional approach of identifying which of the items is the target. The inspiration for this idea came from a study performed by one of the authors (Fox, 2002) in which he acted as receiver in a series of 12 trials without a sender. He felt that he could rule out some of the items in judging sets, but not necessarily identify which of the remaining items was the target.

The present study consisted of 32 ganzfeld trials with novice receivers. In all trials CS acted as sender and JF as experimenter. The study was designated as being exploratory in nature prior to performing the trials and therefore should not be included in meta-analyses. Of the 32 trials, 10 were direct hits (31.25%), the target was ranked second on 7 occasions, third in 13 trials, and fourth in the other 2 (MCE = 8 for all categories).

In addition to awarding a rating of similarity between the items in the judging set and their imagery during the trial period, the participants were asked to rule out any items that they thought could not be the target for the session. Of the 32 trials there were 5 in which 0 items were ruled out, 21 in which 1 was ruled out, 5 in which 2 were ruled out, 1 in which 3 were ruled out, and 0 in which all 4 were ruled out. In the case of 1 of the items being ruled out, participants seemed successful (i.e. the target was only ruled out on 2 of the 21 trials, MCE = 5.25) and the number of direct hits was elevated (7 out of 21, 33%). The low frequencies of ruling out in the other categories make meaningful comments about them difficult to justify.

It is argued that if it is possible to successfully rule out items in the judging set (as suggested from these findings and those of the prior study) then there may be important implications for researchers. For example, if it is possible to successfully rule out one item in the judging set, participants are then left to guess between fewer items (3 instead of 4) and that this would be reflected in the percentage of direct hits observed in studies (33% instead of 25%). If such a process is in operation it would explain why such protocols yield rates of hitting that are only marginally above chance. Furthermore, it is suggested that a change in emphasis from identifying the target to successfully ruling out items that are not the target may have a positive impact on participants.

This study used a digital autoganzfeld system (DigiGanz) to coordinate the trials. Some suggestions are offered to other researchers who are considering using such systems.

INTRODUCTION

For many years the ganzfeld procedure has attracted the attention of researchers in parapsychology. This attention has resulted in some of the most vehement debates in the field, but despite these debates there is still no consensus as to what the database represents; some researchers assert that it demonstrates a genuine psi phenomenon produced in an experimental setting whereas others argue that the apparent support for psi is the result of artefacts of the procedures employed (e.g. Honorton, 1985, Hyman, 1985, Hyman & Honorton, 1986, Bem & Honorton, 1994, Hyman, 1994, Bem, 1994, Milton & Wiseman, 1999, Bem, Palmer & Broughton, 2001).

The importance of the ganzfeld to the discipline is that a body of data has now been amassed over a number of decades, from a number of researchers, to a depth that gives it a unique standing. In addition, it may be seen as holding greater import than other large-scale initiatives because the methodology employed is aimed at studying ESP experimentally whilst using a paradigm that allows the possibility to study the underlying psi processes. This is in contrast to the surveys of researchers such as Gurney, Myers & Podmore (1886), and Louisa Rhine (e.g. Rhine, 1961) who focused on researching reports of events that had already happened, and thus had the associated difficulties of verifying the accounts given, and the experiments of researchers such as JB Rhine (e.g. Rhine, 1937/1950) whose attention was focused on performing a large number of trials in order to see whether people could 'guess' at above-chance levels with little emphasis on the understanding of underlying processes.

The final point to note is that the level of attention that the ganzfeld has received, combined with the scrutiny to which it has been exposed, and the period of time that has elapsed since ganzfeld studies were initiated, has resulted in a paradigm that has been 'improved' over time, and one which now stands on firm foundations. In the three decades of the ganzfeld's existence there have been three phases of development. Initial trials may be described as 'manual' ganzfeld trials as all procedural aspects (e.g. administration of materials, randomisation of targets, and recording of data) were performed manually. This resulted in the potential for human error being able to account for some of the differences between the results that would be expected by chance and those that were observed experimentally (e.g. Hyman, 1985). With these objections in mind, the second phase was initiated through the 'autoganzfeld' procedures which intended to eradicate sources of human error through the use of automated systems. This was achieved through procedures such as the use of computers to randomly select the judging set and target item, and to control video recorders to present these materials to participants (e.g. Berger & Honorton, 1985). Whilst the use of such techniques eliminated many of the possible sources of human bias they did not totally eradicate them; for example Wiseman, Smith, & Kornbrot (1996) suggest the possibility of sensory leakage from the sounds made by the video recorder. Advances in technology have made it possible to enter a third stage of ganzfeld trials using what may be referred to as 'digital autoganzfeld procedures' in which computers are used to control as many aspects of the procedure as possible, and the targets themselves and other media (e.g. white noise) are digitised in order to make it possible for them to be presented through a computer. The aim of such systems is to further reduce the potential for human error, increase the consistency in protocol between trials, and remove as much of the burden of running trials from the experimenter as possible so that they may concentrate on other important issues that have been identified in running studies such as the development of a psi-conducive milieu (cf. Targ, Braud, Stanford, Schlitz, & Honorton, 1991).

In terms of the current status regarding the ganzfeld, it still remains important to collect more data in order to see whether the past effect size may still be maintained. It has already been demonstrated that the level of hitting was not diminished when procedures were tightened from those used in the manual trials to those using autoganzfeld procedures (e.g. Bem & Honorton, 1994, report 35% for ganzfeld studies performed between 1974 and 1981 compared with 32% in the autoganzfeld studies performed between 1983 and 1989). This suggests that the issues addressed by the autoganzfeld systems (e.g. recording errors) were not the source of the higher than chance levels of hitting observed in the manual ganzfeld studies. It would, therefore, be of interest to see whether these levels are retained when using digital autoganzfeld systems.

With the database as it stands, we also have a set of data that requires explanation. Perhaps it does reflect a genuine psi-phenomenon, in which case a set of criteria needs to be set for researchers to accept that this is the case, and perhaps an understanding is needed of the factors that result in some researchers being unwilling to accept the data. Conversely, it may be that the dataset does not represent a genuine psi-phenomenon, in which case many researchers and participants are misinterpreting information, and this needs explaining. In order to achieve these aims it is important to gain as much understanding of the dynamics of the ganzfeld phenomenon as possible, from the manner in which researchers interpret the results of studies, through the manner in which data is collected during trials, to the experiences that all the participants (including experimenters) have during individual trials.

Therefore there is a need for studies to be run that use controlled procedures that take into account past criticisms, gather generic ganzfeld data (e.g. ratings of items in the judging set), explore new avenues (or hunches) through casual observation of sessions, and explore the dynamics of ganzfeld sessions through qualitative analyses of sessions.

This study aimed to tackle these issues through 1) the use of a digital autoganzfeld system (DigiGanz) developed in Liverpool and described previously (e.g. Fox, Smith & Williams, 2002), 2) running a series of 32 trials that could be statistically analysed and compared to the number of hits to the mean chance expectation (MCE) of 8 and prior reported hit-rates, 3) using a novel method of assessing psi 'success' which is described below, and 4) by examining the relationship between a personality variable (tolerance of ambiguity) and the style of mentation produced by the participant. In this paper we deal with all but the last of these. The planned analysis of participants' mentation is beyond the scope of this paper but will be presented elsewhere. Also, at the time of writing, the data relating to participants' ambiguity tolerance has not been analysed, thus its omission from this report is not the result of selective reporting of data. It should also be noted that this study was deemed as being exploratory at the outset and thus the findings should not be included in future meta-analyses.

As mentioned above, one of the aspects examined in this study was the use of a novel method of assessing psi-success. The primary motivation for using this method came from a pilot study performed by one of the authors (JF) in which he took part in 12 trials using a developmental version of the DigiGanz software (Fox, 2002). In the study, JF was his own experimenter and there was no sender. The target was awarded the highest rating in 5 trials, the second highest rating in 6 trials, the third highest rating in 1 trial, and was never awarded the lowest rating. In this series JF had the impression that in many trials it was 'easy' to rule out one or two of the items in the judging set (i.e. state that they were not the target for the session), but that it was not 'easy' to then identify the target from the remaining options. This method is in contrast to the methods of psi 'success' assessment that have been encouraged to date which stipulate that researchers should state which method of analysis they are going to use before the study begins (e.g. whether a 'hit' is defined as a direct hit with 25% chance per trial, a binary hit with 50% chance per trial, or whether a *z*-score is used), and for which the emphasis has been on identifying the target rather than identifying which of the items in the judging set are not the target.

Despite the apparent high level of scoring in JF's exploratory study, it was acknowledged that further data was needed to explore whether using a measure of ruling out options from the judging set would be a useful research measure, and thus it was incorporated into this study. This approach offers a further advantage over other measures as it is participant-centred rather than researcher-centred. This allows the participant to state which item (or items) should be ruled out on the basis of their reaction to the items, rather than allowing the researcher to infer such reactions through, for example, an exceptionally low *z*-score.

METHOD

Design

This study used a digital autoganzfeld system to coordinate ganzfeld trials as described below. There were 32 trials in the study that employed novice receivers. Psi-performance measures considered in this paper are ratings of the items in the judging set (on a scale of 0-100 for each item), whether a direct hit was attained (i.e. the highest rating was given to the target), and receivers were asked to indicate whether there were any items in the judging set that they would have felt surprised to discover were the target for the trial (i.e. whether they would rule out any of the options). For the last measure participants were not required to rule any of the options out, nor were they restricted to only ruling out one of the items.

Participants

One of the authors (JF) acted as experimenter in all 32 trials, the other (CS) acted as the sender in all the trials. The role of receiver was taken by a volunteer in each of the 32 trials. Each volunteer took part in only one trial. All 32 were novice ganzfeld receivers (i.e. they had not acted as receiver in a ganzfeld trial before, though some had acted as a sender, or had been involved in other ESP studies). Participants were recruited on a casual basis through contacts that had been established with people in prior studies. 17 were male, 15 female, with ages ranging from 18.2 years to 58.7 years (mean = 33.5 years, SD = 9.6 years).

Ganzfeld Set-up

The ganzfeld set-up used was based around the DigiGanz software (v1.0.0) running on two networked Apple iMac computers running Mac OS 9. One computer was used to present materials to the sender and was located in the sender's room, the other was used to present materials to the receiver and was located in the receiver's room. The two rooms were in neighbouring buildings and separated by approximately 30m. The computers were networked through a dedicated ethernet link. The system has been described elsewhere (e.g. Fox, Smith & Williams, 2002) and takes care of many procedural aspects of ganzfeld trials such as the selection of the judging set and target for the trial, the presentation of materials, and the recording of the data (including the participants' mentation). In addition, the system used audio streaming software to allow the sender to monitor what the receiver was saying.

Target Materials

The target materials used for this study were 8 sets of 4 items (video clips). The items were filmed by JF as part of the development of the DigiGanz system. All the items were filmed using digital video cameras and edited down to exactly one minute in length. The subject matter for the items were everyday objects, scenes, and activities filmed specifically to be used in ganzfeld trials. In order to achieve this it was intended that the items would be focused on a single theme.

Procedure

Participants were briefed about the procedure prior to taking part in the study and turning up for their trial. This briefing involved explaining the ganzfeld procedure to them and providing an information sheet that summarised this information. The level of explanation offered was varied according to the individual participant as some had specific queries requiring additional information (e.g. about the sensory deprivation and the possibility of claustrophobia), whilst others had either already taken part in ganzfeld trials as senders, or had knowledge of the procedure through general interest in ESP or through studying parapsychology at the college. On the day of the trial the computers in the sender's and receiver's rooms were started up and networked before the arrival of the receiver. The participants were greeted at the receiver's room, asked how they were feeling and whether they had any questions concerning the procedure. If they did, these issues were discussed, otherwise preparations were made for the trial. These were in two parts.

Firstly, participants were asked to fill in a short questionnaire that included details such as age and gender and a tolerance of ambiguity scale (Budner, 1982). Following this the procedure was summarised to them once more with reference being made to the various pieces of equipment (e.g. the computer, microphone, light diffusers) in the room. Following this, the 'tour' of the equipment continued by visiting the sender's room and explaining that CS would be in there throughout the trial, would be viewing the target item, and if appropriate making drawings of aspects that she felt would be pertinent to the successful outcome of a trial. The 'tour' was completed by returning to the receiver's room. This 'tour' was performed in order to ensure that the participants had a good mental picture of the locations of the participants and the nature of the task at hand. Some participants were not given the tour if they had already participated in a prior study as a sender. This decision was made because they were already familiar with the task that the sender would be performing and the location of the room. Therefore it was deemed, in agreement with these participants, that this step would be unnecessary and hinder rather than aid the attainment of the relaxed, informal, atmosphere that was being sought.

The receiver was then prepared for the session by being asked to make themselves comfortable in the reclining chair in which they would be for the session, having the light diffusers (sculpted halved table tennis balls) applied, and the combined microphone and headphone headset placed over their head. As a final step in the preparations the red lamp was illuminated, the main light in the room extinguished, and the red light adjusted to a comfortable level for the receiver.

At this point CS wished the receiver good luck and left the room, locking the door behind her (note that it was possible to unlock the door from inside the room in the case of an emergency). JF stayed with the receiver throughout the session. He then finished entering any trial details for the session into the computer and chatted with the receiver for a few minutes in order to allow CS time to get to her room. As there were no channels of communication into the receiver's room from outside, the amount of time that it would take for CS to get in place was estimated (there were no occasions on which a trial was started before CS was in place). When the receiver was ready the trial was started. This took part in several stages.

Firstly, the receiver and sender were played a short excerpt of white noise in order to ensure that the audio volume levels were comfortable for the session. Following from this there was a five minute relaxation period in which participants were played an extract of music from Cosmic Trigger (Af Storvatten & Lintunnen, 1993). This piece was chosen as it was an ambient piece written as an expression of the artists' experiences following a period of prolonged sensory deprivation in which they spent substantial periods of time in flotation tanks. This music had also been used by one of the authors (JF, unpublished) in a group situation in which reported levels of relaxation were higher following exposure to the music than before. At the end of the relaxation period the music faded out and the white-noise faded in on the receiver's machine indicating the start of the trial period itself.

The trial period lasted for 30 minutes during which time the receiver was continually exposed to the white-noise through their headphones. They had been asked to report any feelings or imagery that they experienced during this period, but at the same time had been reassured that they were not under test and should not allow this aspect of the study to detract from their enjoyment of the ganzfeld. To this end it had been explained to them that through our experience there was a wide variety of reaction to ganzfeld stimulation with some people having experiences from the beginning, whilst for others it may begin later in the session, or that they may experience nothing at all. Similarly it had been explained that some people had vivid imagery, whilst others had vague impressions, and also that some people seemed to be able to report their experiences whilst the session was going, whereas others found doing this too distracting. Finally, it had been made clear to them that if they felt any adverse effects of the stimulation that resulted in them wanting to terminate the session that they were free to do so and that our aim was for people to enjoy the experience of the session. These steps were seen as being especially important in this study due to the fact that the participants were all novices and therefore were being exposed to a novel form of experience.

During the session, the experimenter's role was to ensure that the receiver was not in discomfort, and to transcribe any of the receiver's mentations (audio recordings were being made by the computer, but these written accounts served as a backup in case of computer failure in making the recordings and, more importantly, were to be used as an aide memoire for the receivers following the trial period and during the judging stage).

During the trial period the sender was exposed to the target six times with the start of each exposure being at five-minute intervals. The first exposure was at the beginning of the session. The session, therefore, consisted of six periods of viewing the target item (one minute in length) followed by a four-minute gap. During the trial period the sender could hear the receiver through an audio link and had paper and coloured pencils in order to draw any pictures that she thought were pertinent to the successful outcome of the trial.

The end of the trial period was marked by the fading out of the white-noise. The orientation of the red lamp was then altered so as not to be pointing directly into the eyes of the receiver who was then invited to remove the headset and light diffusers. Following this they were asked to make themselves comfortable by making any adjustments that they wished (e.g. changing the angle of their chair) and were asked how they felt about their experience in the ganzfeld. When they were ready the judging phase of the study was initiated.

This was a two-stage process, the first of which consisted of a review of their mentation during the trial period. This served the dual purpose of reminding participants of their experiences, and allowing them to elaborate upon aspects of what they had reported. The second stage was the judging period itself in which the receiver was shown the four items in the judging set. These were presented full-screen on the computer and in a randomised order determined by the computer. The sender was also presented with the items at the same time. Following from this, miniaturised versions of the items were presented simultaneously on the screen. These were not playing at the time, but were displaying the first frame of each item. Video controllers allowed the receiver to review any part of any of the items.

Having seen the items, the receiver was asked to award ratings to each of the items on a scale of 0 to 100 in terms of the level of similarity between the content of the items and what they considered to be meaningful (target-related) imagery with higher ratings representing higher levels of similarity. Receivers were allowed to give any distribution of scores that they wanted with the one limitation that each of the items had to be awarded a unique rating. They could spend as long as they wanted on this and could also review any of the items or any of their mentation in order to assist in making their decision. Their ratings were entered into the computer by the experimenter through the use of a slider by the side of each item and also recorded manually by the experimenter on a pre-designated section of the participant's information sheet for the session. Once they had decided upon their ratings they were requested to indicate whether there were any items in the judging set that they would be surprised to discover was the target for the session. If any were offered, these were marked on the response sheet by the experimenter.

With the ratings completed, the trial was complete with the exception of the receiver receiving feedback as to the identity of the target. The target was shown on both the receiver's and sender's computer, thus indicating to the sender that she could return to the receiver's room for the end of the session.

The session ended by CS showing the experimenter and receiver the drawings that she had been doing during the trial period, there being a discussion of any interesting aspects of the trial, and the receiver being debriefed and the authors answering any outstanding queries that the participant had.

RESULTS

The results section concentrates on the effect of the manner in which participants ruled out items in the judging set upon their success at identifying the target for the session. This involves two notions of success, 'success' in the traditional sense of being able to identify the target, and 'success' in ruling out items that are not the target.

Overall there were 10 direct hits in the 32 trials (MCE = 8, 31.25%). Of the other trials the target was ranked second on 7 occasions, third in 13 trials, and fourth in the other 2.

Of the 32 trials, there were 5 in which the receiver did not rule out any of the items, 21 in which they ruled out 1 item, 5 in which they ruled out 2 items, 1 in which they ruled out 3 items, and none in which they ruled out all 4 of the items.

Table 1: Summary of responses made by participants to rule out items in the judging set

ruled out	frequency	Probability of success	Successes expected (freq)	Successes observed (freq)	Successes expected (%)	Successes observed (%)
0	5	1.00	5.00	5	100	100
1	21	0.75	15.75	19	75	90
2	5	0.50	2.50	3	50	60
3	1	0.25	0.25	0	25	0
4	0	0.00	0.00	0	0	-

Table 1 provides a summary of one manner in which such data may be considered. Each row summarises outcomes according to the number of items ruled out by participants. For example, the first row deals with trials in which the participants did not rule out any of the items in the judging set. This happened on 5 occasions (the 'frequency' column), the 'Probability of success' is 1.00 representing that by chance alone the target for the session would be expected to be in the remaining (4) items 100% of the time (which by definition must be true), the 'Success expected (freq)' column represents the expected number of successes (i.e. frequency * probability of success), the 'Success expected (%)' column is merely the probability of success expressed as a percentage, and the 'Success observed (%)' is calculated as success observed (freq) divided by the frequency and expressed as a percentage (i.e. multiplied by 100).

As just stated, if none of the items had been ruled out then the target would be expected by chance alone to appear in the remaining items 100% of the time. If one of the items had been ruled out then the target would be expected by chance alone to be present in the remaining items 75% of the time. Similarly, with two of the items ruled out the percentage reduces to 50%, with three of the items ruled out it would become 25%, and finally with all four ruled out the target would be expected to appear in the remaining options 0% of the time (which by definition must be true as the target must have been one of the items that had been ruled out). These expected levels of presence can then be compared to the actual levels of presence. As already implied, two of the options are meaningless as with none or all of the options ruled out the presence of 100% and 0% must occur by definition. This also makes sense logically as the participant in these trials will not have discriminated between the options in the judging set and their likelihood of being the target for the session.

Of the three remaining options we see that the case in which three items were ruled out had only one occurrence and therefore it would be inappropriate to consider observed and expected frequencies. With two items ruled we see 3 successes where 2.5 would be the expected and so this seems to fit the pattern though again with the small number of trials (5) no real significance is attached to this.

The row of most interest is that in which participants ruled out one of the four items in the judging set. Firstly it is worth noting that participants felt confident in ruling one of the items out in the majority of cases (21 of the 32 trials). Here we see that the observed success rate at ruling out an item compared favourably with that expected by chance alone (19 compared to 15.75, or 90% compared to 75%).

Finally we consider the number of direct hits in each category of ruling out. If a participant is not capable of ruling out any options then they are left in the position of guessing between the four items in the judging set with an expected success rate by chance alone of 25%. However, if a participant can successfully rule out one of the items, they are then left to guess between the remaining three items (with an expected 33% success rate). With 2 items successfully ruled out the participant is faced with guessing between the 2 remaining items and therefore has a 50% chance of success. With 3 items ruled out successfully there is only 1 item remaining and therefore they would be expected to have a 100% success rate. The situation of ruling out all 4 should never occur because doing this would not leave any options as a candidate for being the target, which is logically impossible. A summary of the observed frequencies for this study is presented in Table 3.

Table 2: Summary of direct hits in each category if ruling out is successful

ruled out	frequency	expected prob	expected direct hits	observed direct hits
0	5	0.25	1.25	2
1	21	0.33	7.00	7
2	5	0.50	2.50	1
3	1	1.00	1.00	0
4	0	-	-	-

Here we see that in the situation of ruling out 1 of the 4 items in the judging set the expected 1 in 3 success rate was attained (7 of the 21 trials). For the other categories, with none ruled out the frequency is in line with the expected frequency and those for 2 and 3 options ruled out are below expected, but in all 3 of these cases the observed frequencies are small (5, 5, and 1) and so these figures should be treated with caution.

Discussion

This exploratory study has resulted in a number of interesting findings. Firstly in relation to the use of digital autoganzfeld systems, this study has resulted in an encouraging rate of direct hits. The number attained (10) out of the 32 trials equates to a hit rate of 31.25%. Whilst acknowledging that this could easily be accounted for by natural variation on a 1 in 4 hit rate, it should be noted that this is inline with the findings of meta-analyses (e.g. Bem & Honorton, 1994) and thus suggests that there are no intrinsic features of the set-up that eradicates any psi-effect. Thus researchers should feel justified in using such systems in their own research programs.

However, whilst the system successfully coordinated all the trials, audio recordings for some trials were lost following a computer crash, and the audio streaming software did not stream the audio successfully to the sender in all trials (these issues have now been resolved by using newer software). This is mentioned in order to advise others to have backup mechanisms in place (e.g. tape recordings) until they are satisfied that the recordings have been made successfully. Finally, researchers should be aware that such systems may require constant updating. For example, in this study it was not possible to use the system to record the items from the judging set that the receiver ruled out (this aspect was performed using pen and paper). Whilst it would have been possible to record the data digitally additional programming and testing would have been required. Such manual recording of data may result in transcription errors, and therefore leaves at least some of the studies that employ these digital autoganzfeld systems open to one of the criticisms that they have been designed to eradicate.

The findings of greatest consequence in terms of understanding psi-processes are those that relate to the ability of people to rule out items in the judging set. Firstly, in the majority of trials (66%) participants felt confident in ruling out one, and only one, item from the judging set. This points towards participants feeling as though they had gained some information about the target, but not enough to identify the target with confidence. Also, the data for trials in which one item was ruled out is in line with the interpretation that participants were quite successful in using this technique (the target was only ruled out in 2 of the 21 trials, MCE = 5.25) to increase their chances of identifying the actual target (if one item is successfully ruled out the chances of identifying the target by chance alone increase from 1 in 4 to 1 in 3 and there were direct hits in 7 of these 21 trials). The other categories of ruling out (0, 2, 3, and 4 items) had too low frequencies (5, 5, 1, 0) to make meaningful comments about.

These findings point towards a potentially useful new avenue for researchers to explore through considering the nature of the task at hand for the receivers. In traditional approaches to judging in which receivers give ratings to the items in the judging set, the receivers are being asked to interpret the imagery that they have experienced in a manner that allows them to identify the target for the session. This may be a reasonable request to make if the source of imagery were only determined by psi processes, or if the receiver were experienced enough to be able to distinguish between imagery determined by psi and other non-target-related imagery. In reality, however, the source of the imagery that the participants experience is likely to be multi-faceted. Even if some of the imagery is influenced by psi processes, it is likely that much of it is determined by more mundane, day-to-day, factors. Thus the task of the receiver is not only to receive target-related information through psi processes, but also to be able to filter this out from the other imagery that they are experiencing.

In addition, when using novices as participants, even if their imagery is in part determined by psi, it would seem reasonable to assume that they would lack the experience required to distinguish effectively between the target-related material and the 'normal' day-to-day imagery that they may be expected to experience.

If we consider the imagery that the participant experiences throughout the trial period as a whole we may expect the scenario outlined in the paragraphs that follow. This argument is based upon observation and personal experience, and whilst speculative, may provide a framework of understanding some of the findings that are witnessed in ganzfeld and similar studies. Also, the discussion is in terms of broad themes to which there are bound to be exceptions.

If we entertain the notion of psi information existing, it is likely that a receiver's imagery would have four general sources: psi-imagery, imagery relating to experiences that the receiver has had in close temporal proximity to the trial, imagery that has its source in terms of the receiver's hobbies, interests, and generally important themes of life, and finally spontaneous imagery with no apparent source.

If the receiver is then requested to rate the items in the judging set according to their level of correspondence with their imagery, there is little reason to think that the target will be awarded the highest rating even if there is psi-information in the imagery; if the psi imagery is the most dominant in the imagery then a direct hit would be expected, however as the proportion of psi imagery reduces in relation to the other types of imagery mentioned above, the target's ranking will go down if the other imagery matches the other items in the judging set.

With this mechanism in operation, even awarding the target the lowest rating may still be considered a success if the psi information was the least prominent in the imagery. Thus researchers may be expecting too much from their participants when asking them to identify the target for the session, and it may be far more reasonable to ask them whether they can rule any of the options out. This makes sense, especially with participants who lack the experience to be able to classify their imagery into 'target-related' and 'non-target-related' categories, because the experimenter would essentially be saying:

'During the trial period you have experienced a lot of imagery, some of which may be related to the target for the session, and some of which may not be related to the target. Rather than ask you to try to decide which information is target-related, I simply want to know whether there are any items that you think cannot be the target if there was some target-related information in your imagery.'

Taking the approach of ruling out options is not only of interest to the analysis of the outcome of trials. There is also an impact upon the participants in the study and the manner in which they interpret the research project. Using this method allows the participant greater control over the manner in which they express their confidence as to what the target is. Consequently they may feel as though they have not 'failed' at the task if they have not given the target the highest rating (naturally the researchers will discuss possible interpretations of the trial outcome with the participants during the post-trial de-briefing). This may help in terms of the feelings that the participants have towards the researchers and towards participating in future studies.

Not only does this allow participants and researchers to indicate that certain trials could be dismissed as 'void' if they believe that they have insufficient psi-based information to discriminate between items in the judging set, it also provides an explanation as to why one would expect a low effect size even if psi is in operation. Above it has been argued that due to the judging procedures presently used it may be unreasonable to expect the receiver to identify the target. However, it may not be unreasonable to expect there to be trials in which none of their imagery matches one or more items in the judging set. If participants were able to rule out one option then they would be expected to guess the target 33% of the time instead of the 25% expected without ruling the item out. This value of 33% is remarkably similar to the hit-rate reported in meta-analyses.

Naturally, these findings should be treated with caution. Future research comparing this measure to others will shed light on its eventual usefulness. In addition, as noted by one of the reviewers of this paper, although participants seemed successful in the 21 trials in which one item of the judging set was ruled out (with the target only being ruled out twice), in the other 11 trials in which 0, 2, or 3 items were ruled out the target was never given the lowest rating.

Therefore, it is possible that in this study the participants' apparent success in ruling out a single item from the judging set is not due to them being able to do this on a conscious level. Instead it may be that one or more features of the set-up used here (e.g. the nature of the target pool) results in the target rarely been given the lowest rating irrespective of a participant's reaction to the judging set in terms of wishing to rule out items.

References

Af Storvatten, T. & Lintunnen, P. (1993) Cosmic trigger – Polar regions *Silent Records*, Germany.

Bem, D. J. (1994). Response to Hyman. *Psychological Bulletin, 115*(1), 25-27.

Bem, D. J. & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*(1), 4-18.

Bem, D. J., Palmer, J. & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology*, *65*, 207-218.

Berger, R. E. & Honorton, C. (1985). An automated psi ganzfeld testing system. *Research In Parapsychology 1985. Abstracts and Papers from the Twenty-eighth Annual Convesntion of the Parapsychological Association, 1985*. Metuchen: The Scarecrow Press.

Budner, S. (1982). Intolerance of ambiguity as a personality variable. *Journal of Personality*, 30, 29-50.

Fox, J. (2002). The role of introspection in the study of ESP The 26th International Conference of the Society for Psychical Research, Friday 30th August to Sunday 1st September 2002, Manchester Conference Centre, UK.

Fox, J., Smith, M. D. & Williams, C. (2002) Introducing DigiGanz: Describing a digital autoganzfeld system. *The Parapsychological Association 45ht Annual Convention, Paris.*

Gurney, E., Myers, F. W. H., & Podmore, F. (1886) *Phantasms of the living*. London: Trübner.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49*, 51-91.

Hyman, R. (1985). The psi ganzfeld experiment: A critical appraisal. *Journal of Parapsychology, 49*, 3-49.

Hyman, R. (1994). Anomaly or artifact? Comments on Bem and Honorton. *Psychological Bulletin, 115*(1), 19-24.

Hyman, R. & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology, 50*, 351-364.

Milton, J. & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125*(4), 387-391.

Rhine, J. B. (1937/1950). *New frontiers of the mind*. Harmondsworth, Middlesex, Pelican.

Rhine, L. E. (1961). *Hidden channels of the mind.* New York, William Morrow and Company.

Targ, R., Braud, W., Stanford, R., Schlitz, M. & Honorton, C. (1991). Increasing psychic reliability. *Journal of Parapsychology, 55*(March), 59-83.

Wiseman, R., Smith, M. D., & Kornbrot, D. (1996). Assessing possible sender-to-experimenter acoustic leakage in the PRL autoganzfeld. *Journal of Parapsychology, 60*, 97-128.

Address for correspondence: Jezz Fox,
Psychology Department, Liverpool Hope University College, Hope Park, Liverpool L16 9JD, UK.
E-mail: jezz@jezzfox.info

EFFECTS OF FEEDBACK CONTROL ON SLOW CORTICAL POTENTIALS AND RANDOM EVENTS

Thilo Hinterberger[1], Joop M. Houtkooper[2], & Boris Kotchoubey[1]
[1]*Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen*
[2]*Center for Psychobiology and Behavioral Medicine, University of Giessen*

ABSTRACT

As known for more than two decades, humans can learn to achieve self-control over their brain activity such as the polarity of slow cortical potential (SCP) shifts referred to as positivity or negativity. This learning process is supported by the visual feedback of the SCP in a time-locked trial structure. On the other hand, it has been repeatedly shown that humans are able to regulate the random distribution of events, produced by an electronic random event generator (REG) as a psychokinetic effect. This procedure also entails feedback of the distribution of binary random events to the subject; however, contrary to the feedback of SCP, there is no known physical connection from the subject to REG. In both cases, the to-be-controlled function should be changed in one of two alternative directions ("bi-directional control") by use of two different cognitive strategies or states of consciousness. In the present experiment the effect of a self-chosen strategy was examined for the regulation of a feedback signal in a pseudorandomized predefined direction both on the SCP and the REG. Therefore, the system, called "Thought Translation Device" (TTD) provided feedback of the SCP or the REG. This modality was changed every block of 300-400 trials of 4.5 to 5s duration each. The task requirement was assigned pseudorandomly. A correctly produced SCP or REG shift was rewarded with a smily face. The simultaneous recording of the Electroencephalogram (EEG) and the REG signal allowed investigating psychophysiological correlates of psychokinetic effects. This study focused on the differentiation between the two tasks of the REG or SCP signal when feeding back one of them. Four subjects performed in total 3500 trials with SCP feedback and 3500 trials with feedback of the REG. The average feedback value of all trials was analyzed and tested for being different from a baseline value taken before the start of the feedback in each trial using a t-test. The previous findings that human subjects can self-regulate their SCP within the first sessions were replicated. Highly significant SCP control was achieved only by the highly motivated subjects (S 2+S 3) who attended in more than one day. S 2 attained a constant high correct response rate of 70 to 80 % during the second and third training day. It was also shown that a temporary "false feedback" given by the REG did not necessarily disturb the acquisition of SCP self-control. Task-specific lateralization of the SCP (which was not fed back) was not significant. The highly motivated subjects achieved higher t-values in the REG control than S 1 and S 4. When considering both feedback modes together, a significant correlation between pre-defined task requirement and the deviation from chance expectancy produced by the REG ($p=0.02$) was produced. This can be interpreted as a psychokinetic effect. The correlation between task and REG result produced by the system when running the same number of trials was far from significance. Feedback of SCP seemed not to disturb the positive correlation between task and REG. Despite these similarities between the behavior of the REG and the SCPs, the differentiation of the REG signal did not significantly correlate with SCP amplitudes or SCP lateralization.

INTRODUCTION

Slow cortical potentials and their meaning

Slow cortical potentials (SCPs) are potential shifts of the cerebral cortex, which are settled in the frequency range below 1-2 Hz and can persist over several seconds. The SCPs can be measured using the electroencephalography (EEG) or by means of the magnetoencephalography (slow cortical magnetic fields). The amplitudes of the SCP shifts usually vary within a range of 10 to 100 µV RMS and reach a maximum at vertex (electrode position Cz of the international 10/20-system). The fact that the SCPs are not very localized refers to a common activity of expanded neuronal areas. SCPs emerge as synchronous discharge of afferent excitation of the apical dendrites of cortical neurons. These dendrites are located in the upper cortex layer. A negative potential shift (negativity) indicates a lowering of the excitatory threshold and is related to the mobilization of resources for behavioral and cognitive tasks. Positive potential shifts (positivity) can be measured during the execution of cognitive tasks (consumption of resources) or in cognitively inactive states (Birbaumer et al. 1990, Rockstroh et al. 1989). Lutzenberger et al. (1982), for example, showed that subjects could solve arithmetic problems faster after producing cortical negativity. Likewise, response times were shortened if the task was presented during cortical negativity (Rockstroh et al. 1982). The positivity can also result from postsynaptic excitation in deeper cortical layers. Thus firing of the pyramidal cells during cerebral performance can lead to a positivity on the scalp. To sum, negativity represents the mobilization or readiness, positivity represents ongoing cognitive and neural performance or inhibition of neuronal activity.

The relationship between cortical negativity and readiness is best seen in an S1/S2 paradigm which produces the so-called contingent negative variation (CNV): A warning stimulus S1 is presented to a subject and followed by an important "imperative" stimulus S2. Then a cortical negativity appears 300 to 500 ms after S1, which prepares the subject to perform a task after S2 (Walter et al., 1964; Rockstroh et al. 1989, pp. 99). In real life the CNV emerges, for instance,. at a traffic light when expecting the green light and preparing for driving; a positivity can be recorded, however, while the brain is already busy with processing of the stimulus. Although humans are usually not aware of these potential shifts, they can learn to change the amplitudes of the SCP voluntarily into electrically negative or positive direction. This can be achieved by feedback of SCP amplitude changes and positive reinforcement for changes in the correct direction (operant conditioning) (Birbaumer et al. 1981; 1984; 1988; 1992). After having learnt to control the SCPs, humans can also acquire the ability to consciously perceive them (Kotchoubey et al., 2002).

SCP self-control has also been applied to communication by means of a direct interface between brain and computer in completely paralyzed individuals (Birbaumer et al., 1999). These authors have developed a "Thought Translation Device" (TTD) in which self-regulation of slow cortical potentials is used to generate a binary signal. This signal can further be employed to choose letters and words on a computer menu. The TTD has already enabled several completely paralyzed patients diagnosed with amyotrophic lateral sclerosis to communicate solely with their brain potentials (Perelmouter et al. 1999; Hinterberger et al., 2001).

Intention and random event generators

In the last 30 years, studies have been carried out, where the correlations between pre-stated intentions and the output distribution of different kinds of binary random event generators (REGs) have been investigated. A meta-analysis of these studies yielded a highly significant result (Radin & Nelson, 1989). An example of a consistent effort are the studies of Jahn et al. (1997), demonstrating comparable deviations of the mean results from chance expectation in the order of 10-4 bits per bit processed. Although the absolute effect sizes are quite small, these authors showed that the composite effect of a 12-year study exceeded 7 standard deviations, which suggests a very high significance level. The effect does not depend on the distance between subject and the REG device. Even when subjects exerted their efforts at different times from collecting the REG data, the effect sizes were similar. It seems that solely the information about the coherence between an human intention and a (classical) physically independent process is linking these together. The effect vanished when fully deterministic random processes were used, such as the random number generator of a PC, where an algorithm calculates pseudo-randomized numbers. REGs which showed anomalous features were using the thermal noise or quantum noise of electronic components (resistors or diodes) or other physical random processes which are determined by micro-states (e.g. the throwing of dice). This approach led to speculations about the role of consciousness in quantum physics, specifically about the possible influence of an intent observer of a random physical system (Walker, 1979; Houtkooper, 1983; Josephson & Pallikari-Viras, 1991). Although progress towards plausibility has been made (Houtkooper, 2002), a generally accepted explanation for these effects is lacking.

An approach to clarify the interactions between intention and its effect on remote physical processes is to investigate the correlations between the anomalous effects of the REG and the physiological correlates of intention. Such a correlate is produced in every physiological variable which can be self controlled, because self control is an act of intention. Such a variable is the slow cortical potential shifts which can be obtained during the SCP-self control training. For this reason, the self-control of SCPs has been chosen to be compared with the intentional control of REG output in this study.

TTD and REG:

The TTD is a neurophysiological feedback system that can feed back not only different kinds of EEG but any kind of signal such as an REG signal. The TTD program was modified to read in the signals of the REG and handle it simultaneously with the EEG signal processing. Thus the REG signal could be fed back to the subject and its self-control could be trained, if possible. The intention was to collect physiological variables which significantly correlate with REG control, and thereby to clarify the issue of physiological processes possibly mediating psychokinetic (PK) effects. An advanced application of such physiological correlates could (reversely) raise the question of whether it is possible to facilitate PK effects by feeding back and training those variables. This study provides information concerning the following questions:

1. Is there a correlation between the required intention and the result of the REG (the PK effect)?
2. Can this PK-effect be improved by the feedback of the REG result?
3. Is there also a significant PK-effect in case of feeding back the SCP?
4. How does the attempt to control the REG correlate with the EEG (especially the SCP) of the subject?

METHODS

Feedback and SCP self control

The feedback of slow cortical potentials is provided in a setup shown in Figure 1. The EEG-signal was conducted by means of Ag/AgCl electrodes placed at the vertex of the head and amplified in the EEG amplifier that was connected to the TTD via an A/D converter. The impedances of the electrodes were below 5 kOhms. One central electrode (Cz; international 10/20-system) served as the active electrode for the feedback. The potential at Cz was referred to the mastoids at position A1 and A2. Since it has been shown that also lateralized SCP self control can be learned (Kotchoubey et al., 1996a) the signals at the positions C3 and C4, i.e. over the left and right motor cortex were recorded too. The EEG was sampled with 256 S/s and recorded with a frequency range from 0.01 Hz to 40 Hz. The low cut-off frequency of 0.01 Hz was important since very slow EEG components were fed back; therefore a time constant of more than 10 s was required. To control and correct artifacts caused by vertical eye movements and blinks, two additional electrodes above and below one eye were attached (vertical electro-oculogram, vEOG). The feedback-signal was calculated on-line using the mean of the channels Cz-A1 and Cz-A2 corrected with the vEOG. The algorithm for the artifact correction is described in Kotchoubey et al. (1996b; 1997). Artifacts larger than 1 mV and EEG fluctuations larger than 200 µV led to the cancellation of a current trial (invalid data). The SCP feedback-signal was generated from low-pass filtering of the artifact-corrected EEG using a sliding averaging window of 500 ms. Thus frequencies above 2 Hz were filtered out, leading to a smooth movement of the feedback cursor which was updated 16 times per second. This cursor was symbolized by a yellow circle (ball) whose vertical position reflected the actual SCP value. Cortical positivity moved the cursor downwards whereas cortical negativity lead to an upward movement. The screen also showed the randomly alternating task requirements (negativity versus positivity) by two rectangles ("goals"), displayed at the upper or lower edge of the screen. The illumination of a goal indicated the direction to which the cursor should be moved. Successes were reinforced with a "smiley face". The operator was sitting in another room where he could control the experiment (for details see Hinterberger (1999)).

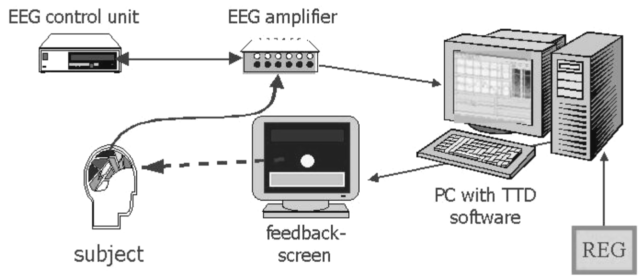


Figure 1: Experimental setup: An eight channel EEG-amplifier is connected to a personal computer with the TTD-software. The random event generator is connected to the serial port. Both data types can be fed back to the subject as the movement of a yellow cursor on a second monitor. The required intention is either to move the ball upwards or downwards.

Feedback of random event generator signals

The REG produced a sequence of binary numbers. The proportion of ones and zeros was equally distributed and was supposed to be psychokinetically influenced. The binary-coded data were transmitted into the PC with a data transmission rate of 9600 Baud. Thus approx. 3850 ones and approx. 3850 zeros were generated per second and transferred. In the present experiment, the difference between the number of ones and zeros in a certain time interval served as the REG signal amplitude. Because the TTD was already successfully used as SCP training and feedback software, only small modifications were necessary to use it for REG feedback. The REG data were read and handled as a separate EEG channel in the software. There was no electronic interference between the analogue REG signal and the EEG signal, because the REG data were read in as an already digitized signal through the serial port whereas the EEG was digitized by an A/D-board in the computer. This zero correlation between the two signals is of great importance since previous data (Hinterberger, 1999) indicate that task related differences in SCP may be very large, whereas task related differences in the REG signal, even if significant, are expected to be very small. Therefore, even a weak correlation between the two channels might lead to REG signal changes that would be erroneously interpreted as a PK effect. A digital switching mechanism inside the REG guaranteed that possible influences of slow waves on the analogue noise of the REG cannot disturb the digital random distribution.

To make the feedback of the REG signal similar to the feedback of SCPs, the same paradigm and the same calculation method was applied to the difference between ones and zeros as to the EEG. The screen for the feedback signal also was kept identical. The vertical ball movement reflected the difference between ones [1] and zeros [0] within the time interval of the last 500 ms. To achieve an upward movement more ones than zeros had to be produced, and a downward movement required more zeros than ones. Each feedback value $F_i(t)$, calculated from the last t_{FB}=500 ms consisted of 3850 random bits:

(1) $$F_i(t) = \sum_{tn=t-t_{FB}}^{t} ([1]-[0])_{tn} = F_i$$

The feedback was also updated each 1/16 s, leading per trial of 4.5 s duration to 72 feedback values. These are indicated with the time indices i=1..72. Regardless of the signal used for feedback (SCP or REG), a data file was created containing both signals for off-line analysis. Thus SCP shifts and other EEG components could be investigated while the subject was trying to control the REG and vice versa.

The feedback paradigm

The feedback training of SCP and REG signals was conducted in a sequence of individual trials with no intertrial intervals. 100 trials constituted a run after which a short resting period was permitted. A trial lasted 4.5 to 5 seconds and consisted of two time intervals:

1. a preparatory interval of 2 s duration. The subject received the information about the following task by illumination of the upper or lower rectangle on the screen. At the end of the preparatory interval, the current signal level was set as baseline level, which corresponded to the vertical center position of the ball as starting position for the following feedback. This interval was followed by

2. the feedback interval of a duration between 2.5 to 3 seconds. Here the subject received feedback over the SCP or the REG and had to move the ball towards the rectangle that was still illuminated during this interval, too. The movement of the ball was a linear function of the feedback value F_i. E.g., if the upper goal was lit, the ball should be moved upwards during the 2.5 to 3 s feedback time. This was achieved by producing cortical negativity in the case of SCP feedback or by producing more ones than zeros in the case of REG feedback. The reverse was true for the lower goal. If the average of all ball positions was in the correct half of the screen, the subject received a positive reinforcement by a smiling face ('smiley '), which was presented during the final 500 ms of the trial after the feedback.

Experimental design

The two criteria for selection of subjects were the ability to concentrate during the experiment and the belief in parapsychological phenomena. Therefore, subjects were selected who had experience in transcendental meditation (TM) for many years. Four subjects (two female and two male) ranging in age from 40 to 60 years took part at the study. The subjects were seated in a comfortable chair in a small isolated, electromagnetically shielded room. Each subject participated in the experiment on at least one training day comprising 1000 trials. The subjects were instructed that there were two kinds of feedback signals (i.e. A or B). They were informed when there was a switch between signals A and B, but they had no knowledge about the nature of these signals (i.e., that the REG signal served as A and the SCP signal, as B). Beforehand, they only were informed that this was a parapsychological experiment.

Table 1: Each of the four subjects attended the study in one to three days, depending on his/her own motivation. One training day comprised about 1000 feedback trials of either the REG or the SCP. The modality was altered twice resulting in three trial blocks per day.

training day	S 1 (m) feedback signal, trials	S 2 (m) feedback signal, trials	S 3 (f) feedback signal, trials	S 4 (f) feedback signal, trials
1st day	REG, 300 trials	REG, 300 trials	SCP, 300 trials	SCP, 300 trials
	SCP, 400 trials	SCP, 400 trials	REG, 400 trials	REG, 300 trials
	REG, 300 trials	REG, 300 trials	SCP, 300 trials	SCP, 200 trials
2nd day	-	SCP, 300 trials	REG, 300 trials	-
	-	REG, 400 trials	SCP, 500 trials	-
	-	SCP, 300 trials	REG, 300 trials	-
3rd day	-	REG, 300 trials	-	-
	-	SCP, 400 trials	-	-
	-	REG, 300 trials	-	-

Table 1 illustrates the experimental schedule leading to almost 7000 trials over all. With this setup the following questions and interactions can be explored:

1. Analysis of the task specific SCP-shifts:

a) Can subjects learn to self-control their SCP?

b) Is this learning process critically dependent on the SCP feedback?

2. Analysis of the task specific REG results:
a) Can subjects significantly influence the REG result in the desired direction?
b) If yes, does this effect depend on REG feedback?

Analysis of task specific SCP shifts:

To analyze task specific SCP-shifts, all trials were averaged in the time domain, separately for the positivity task and the negativity task, leading to two average EEG waveforms for a trial. The ability to self-control the SCP can be seen in the amplitude difference between the two tasks during the feedback interval. This difference, called SCP differentiation (see figure 3), was calculated as difference between the mean amplitude during the second half of the feedback interval. A t-test was applied to calculate the significance of this differentiation (see below, Eq. 5).

Another measure for the ability of SCP self-control is the correct response rate. It is the percentage of correctly classified cursor movement responses. During on-line training the classification algorithm calculated a response for each trial as the integral of all cursor positions during the feedback. A correct response was counted when the sign of this integral matched with the task requirement (positivity or negativity). An off-line classification using a discriminant analysis (Hinterberger, 1999) instead of a simple integral can lead to higher correct response rates but needs information of previous runs. However, this method supplies more precise information about the ability to produce two different signals and thus was used off-line. As there was a strong correlation between off-line correct response rate and SCP differentiation (r=0.92), the SCP-differentiation can be regarded as a satisfying measure for performance.

Analysis of task specific REG results:

As already mentioned, each feedback value F_i was the difference between approx. 1925 ones and 1925 zeros. The standard deviation σ_i for F_i can then be estimated for equally distributed values as

(2) $$\sigma_i = 2 * \sqrt{\frac{[1]_i + [0]_i}{2} \Big/ 2} \approx 62 .$$

The assumption with these formulas is that the number of ones is binomially distributed with p=1/2. For our analysis the standard deviation was calculated using the actual REG numbers. The uncertainty $\sigma_{g,i}$ for the mean G_i over N_g trials with $N_g=N^{(1)}+N^{(0)}$ is

(3) $$\sigma_{g,i} = \sigma_i \Big/ \sqrt{2N_g} \quad \text{for} \quad G_i = \sum_{N_g} F_i .$$

can be estimated by

(4) $$\hat{\sigma}_i = \sqrt{\frac{\sum_{N^{(1)}} (F_i^{(1)} - G_i^{(1)})^2 + \sum_{N^{(0)}} (F_i^{(0)} - G_i^{(0)})^2}{(N^{(1)} - 1) + (N^{(0)} - 1)}}$$

The means $G_i^{(1)}$ resp. $G_i^{(0)}$ of the $F_i^{(1)}$ resp. $F_i^{(1)}$ over $N^{(1)}$ resp. $N^{(0)}$ trials were calculated separately for each task. The significance of REG control was assessed by means of a t-test (Bortz, 1999).

(5) $$t_i = \frac{G_i^{(1)} - G_i^{(0)}}{\hat{\sigma}_i \cdot \sqrt{\left(\frac{1}{N^{(1)}} + \frac{1}{N^{(0)}} \right)}} ,$$

where $G_i^{(1)}$ and $G_i^{(0)}$ - the means for the tasks to produce more ones than zeros or more zeros than ones, respectively; $\hat{\sigma}_i$ – the corresponding standard deviation; $N^{(1)}$ and $N^{(0)}$ – the number of trials of each task, in which the corresponding $\hat{\sigma}_i$ and G_i were calculated.

RESULTS

Analysis of task specific SCP shifts:

Subjects S 1 and S 4 were not very motivated and took part in the first training day only. With the SCP feedback, S 1 could achieve an average SCP differentiation of 2 µV resulting in a maximum correct response rate of about 60 %, which is significantly better than 50 % expected by chance ($t=4.0$, $N=400$, $p<0.001$). The correct response rate for SCP during REG feedback did not differ from chance. S 4 learned to self-control her SCP with SCP feedback and attained 60 and 67 % of correct responses ($t= 5.92$, $p<0.001$) in the 1st and 2nd block, respectively (see Figure 2).

S 2 was highly motivated and achieved remarkable self-control of SCP. SCP shifts and the correct response rates were highly significant ($p<0.001$) in every block. Also the feedback of the REG signal had no negative influence on the SCP self-regulation. The subject produced SCP differentiations of more than 20 µV and achieved 70 to 80 % correct responses except for the first training block, in which the REG and not the SCP was presented as feedback signal. Figure 3 shows no considerable difference between the waveforms obtained during SCP feedback and REG feedback. In trials with required positivity, S 2 produced a negative potential shift during the preparatory interval which caused a positivity relative to that negative baseline level in the feedback interval.

S 3 also attained significant control of her SCP-shifts ($t=6.1$ and 6.7 in the first and second SCP block; both $p<0.001$). However, she changed her strategy during REG feedback leading to a highly significant SCP differentiation (all REG blocks: $t=-7.1$, $p<0.001$) in the wrong direction. After repeated REG feedback, with SCP feedback in the runs 14-16, she could not obtain SCP control any longer (s. Figure 3). These runs were also contaminated with eye movement artifacts.

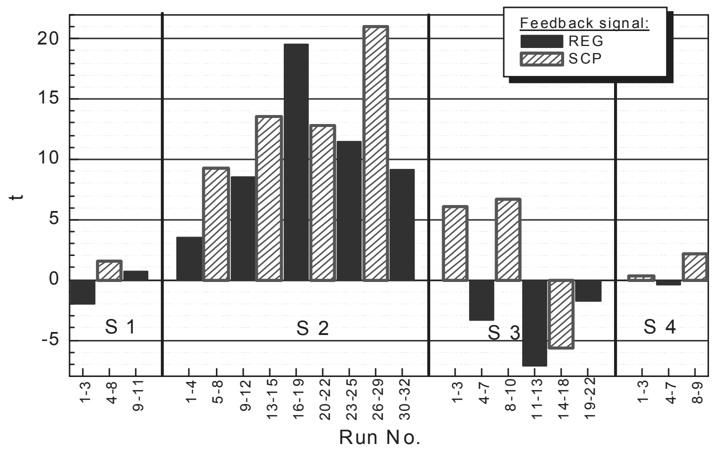


Figure 2: Significance (t-values) of the deviation from chance for SCP differentiation during SCP feedback (shaded bars) and REG feedback (filled bars). Blocks of 300 to 500 trials, conducted under equal conditions, were averaged for each of the four subjects separately and analyzed. Positive t-values stand for correct differentiation, negative t-values indicate discrimination in the wrong direction (e.g., larger negative SCP shifts when positive shifts were required).

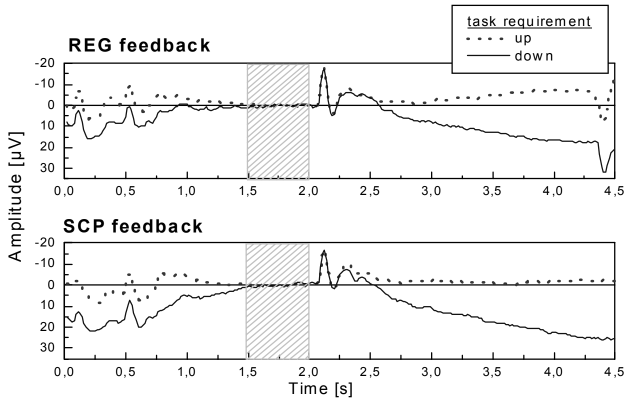


Figure 3: Averaged (over all trials) SCP waveforms of subject 2, separately for two task requirements, during REG feedback (top) and SCP feedback (bottom). The shaded area indicates the baseline period.

Lateralization: Although the feedback of left-right SCP-differences was not presented, in some blocks subjects showed significant lateralized SCP shifts (see Figure 4). Mostly, when a downward movement of the cursor (cortical positivity during the SCP feedback) was required, this positivity was larger over the right hemisphere as compared with the left hemisphere during the feedback interval. In addition, subject S 2 produced a significant right hemispherical negativity during the required upward cursor movement during the first training day,. This effect vanished in the following training days. The correlation between the differentiation of the central SCPs and the lateralised SCPs was not significant (r=0.15, p=0.22, N=66).

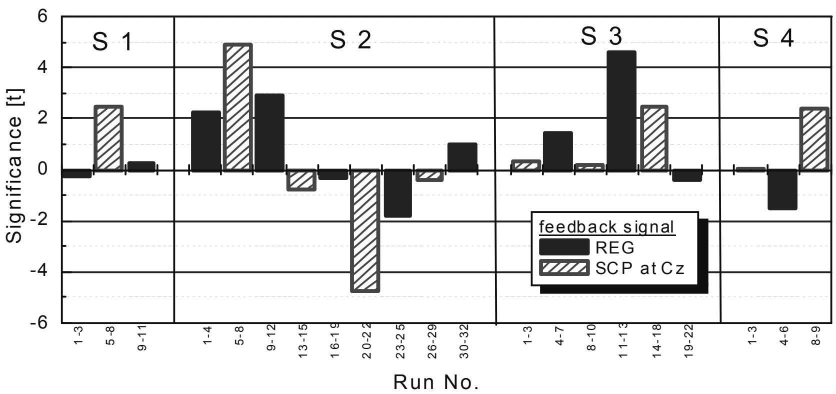


Figure 4: SCP lateralization measured at electrode positions C4-C3 during feedback of SCP and the REG. Positive values indicate that the left hemisphere is more positive while the cursor should be moved upwards, whereas negative values indicate a left positivity while the cursor should be moved downwards.

Task specific REG results:

Figure 5 shows the t-values of the deviation of the achieved REG results from chance, which would be an equal number of ones and zeros. The REG changes were analyzed like the SCP changes reported above and were taken as the difference between the currently measured REG signal and the baseline level (i.e. the mean of the 500 ms before the FB-interval starts).

Subject S 1 started with REG feedback and did not attain a significant influence on the REG (t=0.62, p=0.27, Ng=1007; one tailed t-test). Also S 4 showed no significant REG results (t=-0.66, p=0.75, Ng=507).

In contrast, S 2 achieved a high correlation between task requirement and REG result in the first block with REG feedback (t=2.4, p<0.01). The following block with SCP feedback yielded a sizeable negative result (t=-2.0, p=0.98). A highly significant result (t=3.05, p<0.01) was then attained in the 6th block with SCP feedback. Although 7 of 9 blocks yielded positive t-values, S 2's data across all trials did not reach the .05 significance level (t=1.48, p=0.07, Ng=2021).

We test one-tailed in the direction of intention. Other ways of looking at the data are the REG data with REG feedback: 9 out of 11 'blocks' score in the positive direction; the sign test gives p = .09. The two negative deviations occur with the two subjects who achieved less SCP self-control and who were less motivated to continue. The overall 16 positive out of 20 blocks has a p = .006 according to the sign test (Siegel, 1956).

Subject S 3 achieved positive t-values in all six training blocks with an overall significant result of t=1.72 (p=0.04, Ng=2115).

Across subjects, blocks with REG feedback did not lead to a significant result t=1.20, p=0.12) but blocks with SCP feedback did (t=1.69, p=0.046). Overall, positive t-values were obtained in 16 of 20 blocks, resulting in a significant overall t=2.09 (N=6951, p=0.018). The mean correct response rate did not differ significantly from chance. This significant correlation between REG feedback signal (which is calculated as the difference between the REG signal during the feedback interval referenced to the baseline) and the task requirement could be mainly attributed to the REG behavior during the baseline ($t_{baseline}(6950)=1.82$, p=0.03) while the REG result during the feedback interval did not significantly contribute to this effect.

The differentiation of the REG was found to correlate neither with SCP amplitudes (across runs r=-0.063, p=0.62, N=66), nor with the SCP lateralization (r=-0.044, p=0.73, N=66).

To exclude that the system itself produces systematic correlations between the task (determined by the software random number generator of the PC) and the REG result 7000 trials were collected without a subject watching the feedback. The overall t(7000)=.17 (p=.43) was far from significance.

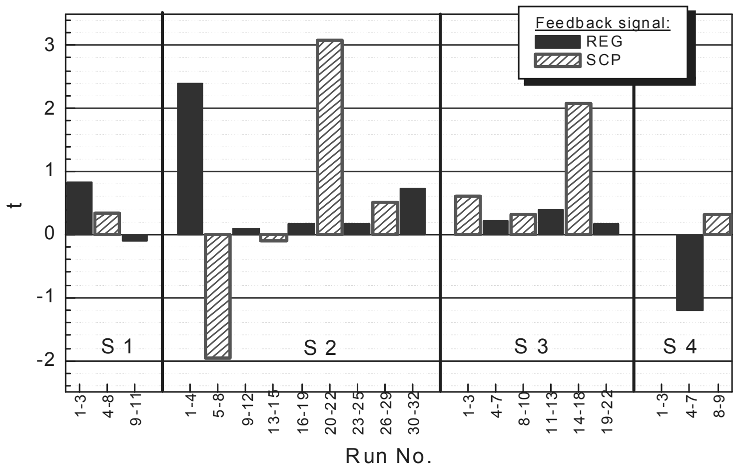


Figure 5: Significance (t-values) of the deviation from chance of the REG result while feedback of SCP (filled bars) and of the REG result (shaded bars) was presented. Blocks of 300 to 500 trials, conducted under equal conditions, were averaged for each of the four subjects separately and analyzed.

DISCUSSION

The ability of humans to learn to self-control their SCPs is already well known (Birbaumer, 1984; Rockstroh et al., 1989). Surprisingly, however, three of four subjects participating in the present experiment achieved this self-control already in the first runs. Moreover, the data of S 2 show that this acquired ability is not necessarily disturbed or impaired when a random signal (such as the REG) is presented as the feedback. This stability indicates that the strategy may be more important than the feedback signal. Such an interpretation is in line with the self-control theory of Lacroix (1981) who suggested that during operant conditioning of bodily functions, subjects frequently (at least on the first stages of training) select a strategy from their already existent cognitive-behavioral repertoire and keep this strategy as long as it does not result in a clear failure. In the present data, such strategy (particularly employed by S 2) was the development of a negative SCP shift during the baseline interval to support subsequent positivity. During the baseline, waiting for the stimulus indicating the onset of the feedback interval served as a condition in which a negativity (i.e., the contingent negative variation (CNV), see Walter et al., 1964) could easily be produced (see also Brunia, 1993).

This stability of self-regulating strategy can also be a particular trait of S 2, because in S 3, in contrast, the presentation of the REG signal as feedback did deteriorate the already acquired SCP control. The present data do not allow to specify factors which might determine the outcome of this conflict.

Again in line with the data of the literature (Radin & Nelson, 1989), a very small but significant ($p<.02$) effect of task requirement on the REG result, indicates that the technique employed in the present study allows to control an electronic REG in the desired way. When reformulating the hypothesis to test for an interaction between task (which was connected to a certain cognitive strategy) and REG result in any direction, the two-tailed t-test still remains significant ($p=.04$). Despite the stability of this PK effect, there is no generally accepted explanation for it to date. Tentatively, however, we might suggest that such putative factors as precognition or extra-sensory perception could not play a role here because the task was not chosen by the subject but generated by the algorithm long before the start of the experiment. Factually, a significant covariation was found between the task generated by the software random number generator and the REG. However, an explanation for such a correlation is hardly thinkable. A non-significant correlation between task and the REG feedback signal was found when the same number of trials were run without a human interaction.

Notably, the significant influence on the REG in the desired direction was obtained when SCP feedback was presented, but not when REG feedback was presented. This result may appear surprising but it is, again, in agreement with the literature data demonstrating considerable effects of motivation and self-confidence on PK effects (Stanford, 1977). During SCP training, subjects successfully produced the required shifts of the cortical potential and were consistently rewarded for their performance. They were not aware of the principal difference between the two kinds of self-control, that is, that in one case there was a physical connection between their brain and the controlled cursor on the screen, and in the other case there was no. Rather, they just knew that they were involved in two different, but equally challenging, self-regulation tasks. They also experienced themselves being quite successful in one of these tasks. This success may have encouraged subjects thereby resulting in a higher PK effect. This motivational explanation remains speculative at present. In future PK experiments using a larger sample of subjects the motivation should be directly controlled by means of a psychological instrument.

Several parapsychological studies reported, further, that PK effects are the largest at the very beginning of an experiment and then decrease across trials (e.g., Irwin, 1994). This appears to hold true for our subjects S 1 and S 2 who started with REG feedback and attained good results in those first runs that they could not repeat in further runs. In contrast, S 3 and S 4 who started with SCP feedback, achieved better results in SCP runs later in the course of training. It appears, furthermore, that while the PK effect decreased across blocks with REG feedback, it increased in blocks with SCP feedback, which is also compatible with the motivational explanation given above.

An alternative to this motivational explanation may be, of course, some specific effect of SCP changes, based on the fact that these changes at the central electrode are related to regulation of excitability of large cortical regions as mentioned in the introduction.

From the point of view of SCP training, REG feedback can be referred to as 'false feedback'. Particularly, the small size of the PK effect makes the task very frustrating, thus the subjects can resignate and their resignation can generalize to the easier SCP regulation task. A better result might be obtained with mixed feedback, which would contain partly SCP and partly REG trials to support some level of success and to avoid frustration. Alternating pure SCP with mixed feedback (announcing it as a more difficult task) or mixed feedback alone might therefore be explored.

In further studies, the presently used TTD can be extended to a generalized program to feed back various physiological parameters correlated to PK-effects. Thereby self-regulation of all these parameters can be trained to check their influence on the PK-effect. Such a training program might then enable people to develop their PK abilities.

ACKNOWLEDGEMENTS

We thank our assistant Slavica von Hartlieb for help in the measurements. We also thank the Institut für Grenzgebiete der Psychologie und Psychohygiene, Freiburg i.Br., Germany for the financial support.

REFERENCES

Birbaumer, N. (1984). Operant control of slow brain potentials: a tool in the investigation of the potential's meaning and its relation to attentional dysfunction. Pages 227-239 in T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer, eds. *Self-Regulation of the Brain and Behaviour*. Springer-Verlag, Berlin.

Birbaumer, N., Elbert, T., Rockstroh B. & Lutzenberger, W. (1981). Biofeedback of event-related slow potentials of the brain. *International Journal of Psychology 16*, 389-415.

Birbaumer, N., Elbert, T., Canavan, A. G. M. & Rockstroh, B. (1990). Slow potentials of the cerebral cortex and behavior. *Physiological Reviews 70*, 1-41.

Birbaumer, N., Ghanayim N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E. & Flor. H. (1999). A spelling device for the paralysed. *Nature 398*, 297-98.

Birbaumer, N., Lang, P. J., Elbert, T., Lutzenberger, W. & Rockstroh, B. (1988). Slow brain potentials, imagery and hemispheric differences. *International Journal of Neuroscience* 39.

Birbaumer, N., Roberts, L. E., Lutzenberger, W., Rockstroh, B., & Elbert, T. (1992). Area-specific self-regulation of slow cortical potentials on the sagittal midline and its effects on behavior. *Electroencephalography and Clinical Neurophysiology 84*, 351-361.

Bortz, J. (1999). *Statistik*. Springer-Verlag, Berlin, p. 138.

Brunia, C. H. M. (1993). Waiting in readiness: Gating in attention and motor preparation. *Psychophysiology 30*, 327-339.

Hinterberger, T., Kaiser, J., Kübler, A., Neumann, N. & Birbaumer, N. (2001). The Thought Translation Device and its Applications to the Completely Paralyzed. In Diebner, Druckrey &Weibel: *Sciences of the Interfaces*. Genista-Verlag Tübingen.

Hinterberger, T. (1999). Entwicklung und Optimierung eines Gehirn-Computer-Interfaces mit langsamen Hirnpotentialen. *Dissertation in der Fakultät für Physik an der Eberhard-Karls-Universität Tübingen*, Schwäbische Verlagsgesellschaft: ISBN 3-88466-177-9.

Houtkooper, J.M. (1983) *Observational theory: A research program for paranormal phenomena*. Lisse, The Netherlands: Swets & Zeitlinger.

Houtkooper, J.M. (2002) Arguing for an observational theory of paranormal phenomena. *Journal of Scientific Exploration, 16*, 171-185.

Irwin, H.J. (1994) *An introduction to parapsychology (2nd edition)*. Jefferson, N.C.: McFarland.

Jahn, R.G., Dunne, B.J., Nelson, R.D., Dobyns, Y.H. & Bradish, G.J. (1997). Correlations of Random Binary Sequences with Pre-Stated Operator Intention: A Review of a 12-Year Program. *Journal of Scientific Exploration*, *11*, 345-367.

Josephson, B.D. & Pallikari-Viras, F. (1991) Biological Utilisation of Quantum NonLocality. *Foundations of Physics, 21*, 197-207.

Kotchoubey, B., Kübler, A., Strehl, U., Flor, H. & Birbaumer, N. (2002). Can Humans Perceive Their Brain States. *Consciousness & Cognition 11,* 98-113.

Kotchoubey, B., Schleichert, H., Lutzenberger, W, Anokhin, A. P. & Birbaumer, N. (1996a). Self-regulation of interhemispheric asymmetry in humans. *Neuroscience Letters, 214*, 91-94.

Kotchoubey, B., Schleichert, H., Lutzenberger, W. & Birbaumer, N. (1997). A New Method for Self-Regulation of Slow Cortical Potentials in a Timed Paradigm. *Applied Psychophysiology and Biofeedback, 22*, 77-93.

Kotchoubey, B., Schneider, D., Schleichert, H., Strehl, U., Uhlmann, C., Blankenhorn, V., Fröscher, W. & Birbaumer, N. (1996b). Self-regulation of slow cortical potentials in epilepsy: A retrial with analysis of influencing factors. *Epilepsy Research, 25*, 269-276.

Lacroix, J. M. (1981). The acquisition of autonomic control through biofeedback: A case against an afferent process and a two-process alternative. *Psychophysiology, 18*, 573-587.

Lutzenberger, W., Elbert, T., Rockstroh, B. & Birbaumer, N. (1982). Biofeedback produced slow brain potentials and task performance. *Biological Psychology, 14*, 99-111.

Perelmouter, J., Kotchoubey, B., Kübler, A. & Birbaumer, N. (1999). Language support program for thought-translation devices. *Automedica, 18*, 67-84.

Radin, D.I. & Nelson, R.D. (1989). Evidence for Consciousness-Related Anomalies in Random Physical Systems. *Foundations of Physics, 19,* 1499-1514.

Rockstroh, B., Elbert, T., Canavan, A., Lutzenberger, W. & Birbaumer, N. (1989). *Slow Cortical Potentials and Behavior*. Urban & Schwarzenberg, Baltimore.

Rockstroh, B., Elbert, T., Lutzenberger, W. & Birbaumer, N. (1982). The Effects of Slow Cortical Potentials on Response Speed. *Psychophysiology, 19*, 211-217.

Siegel, S. (1956). Nonparametric Methods for the Behavioral Sciences. New York: McGraw-Hill, pp.68-75.

Stanford, R.G. (1977). Experimental Psychokinesis: A Review from Diverse Perspectives. In B.B. Wolman, Ed.: *Handbook of Parapsychology*. New York: Van Nostrand Reinhold, 324-381.

Walker, E.H. (1979). The Quantum Theory of Psi Phenomena. *Psychoenergetic Systems, 3*, 259-299.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C. & Winter, A. L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 380-384.

Address for correspondence: Thilo Hinterberger,
Institute of Medical Psychology and Behavioral
Neurobiology, Eberhard-Karls-University Tübingen,
Gartenstr. 29, D-72074 Tübingen.
E-mail: thilo.hinterberger@uni-tuebingen.de

EXPLORING VOLITIONAL STRATEGIES IN THE MIND-MACHINE INTERACTION REPLICATION

Joop M. Houtkooper

Center for Psychobiology and Behavioral Medicine, University of Giessen

ABSTRACT

Volitional strategies in PK experiments have been reported by experimenters acting as their own subjects, mainly in the middle decades of the 20th century, but also by the PEAR group. A study by Gissurarson, in which he obtained volitional strategies from the participants after the PK task, put this on a broader basis. His work finds a continuation in the present experiment, conducted as part of the Mind Machine Interaction (MMI) collaboration.

The present study involved 74 participants who took part in 271 PK (or MMI-) sessions of about an hour. A session consisted of 30 short or 3 long runs, balanced between the intentions "High", "Low" and "Baseline". The overall PK scoring was in the intended direction, that is, the difference of the RNG output between High and Low intentions was positive, but fell short of significance (p=.24). However, the salience (U-shape) pattern during the sessions revealed some indication of PK (z=2.35, p<.02).

A novel aspect was that the participants recorded their volitional strategy before each PK-run, from a list of 5 choices derived from the inventory by Gissurarson, namely "Imagery", "Relaxed", "Confidence", "Resonance" and "Guessing", as well as the option to employ their own strategy. The consensus in the literature, that a strategy involving "effortless effort" or "resonance", would be the most successful, could be confirmed (p=.03).

Furthermore, it appears that the choice of a volitional strategy is related to several psychological variables. Participants who rated themselves high on Luckiness, chose more often the resonance strategy or their own strategy. They obtained also significant above chance PK scores on runs in which they used the resonance strategy (p<.03), or their own strategies (p<.03). On the other hand, those with a high ability to visualize chose less often their own strategy, but scored significantly on those runs (p<.01). Participants who scored high on the Conscientiousness scale of the NEO-FFI chose the resonance less often, but scored very significantly (p<.001) on the runs in which the resonance strategy was employed.

An item in the questionnaire relevant for absorption showed a positive correlation with the preference for the resonance strategy. Therefore, choosing the resonance strategy more often in the experiment is related to being more absorbed generally. There are indications that successful PK performance depends more on the momentary choice of a volitional strategy and the concomitant state of mind than on the overall preference of a person to choose certain strategies.

INTRODUCTION

One of the greatly puzzling, but inconspicious things in PK experiments is the fact that almost everybody, when asked to mentally influence a random process, sits down without much ado and does something. But what?

However, the first question about this should be: How it is possible that virtually every participant in these experiments has some idea about what to do when a mental influence is called for? The second question is: Even if they do individually very different things, what do they do and what brain states might correspond with these different modes of mental effort?

The present study is concerned with part of the second question, that is, what "volitional strategies" are employed by participants in a PK experiment. The study by Gissurarson (1997) has led to an inventory of volitional strategies. A number of the strategies he has found were used in the present experiment in a prospective setting, instead of having the participants describe afterwards the volitional strategies they used.

About the description of beneficial psychological conditions in a PK experiment, the early experimenters, who often acted as their own subjects, have reported some introspective and anecdotal material. For instance, Thouless (1951) and Forwald (1952, 1969) have been such experimenters. In this tradition, of the experimenters taking part in their own experiments, also belongs the Princeton Engineering Anomalies Research (PEAR) group. Indeed, they have reported some of the subjective findings by the participants in their experiments (Jahn & Dunne, 1987, p.142). Consistent with these reports is a study by Schmidt (1997) in which he observed psi-missing under a variety of stressful circumstances. There is some similarity in these reports, namely that there is a state of mind, described as "effortless effort" or "resonance" with the apparatus, which is associated with success in a PK task.

There have been other indications, that being relaxed works better than intense striving for an effect. Stanford (1977) has reviewed a broader range of psychological factors which might be associated with success in PK experiments, while Irwin (1994) more specifically has discussed states of consciousness associated with success in PK tasks, pointing to an unforced state of absorption in the goal of the task facilitating PK performance.

Another approach to successful PK might be the study of spontaneous cases, such as those described by Irwin (1994, p.132-3) and the poltergeist cases. However, if we want to apply the knowledge in this area to an experimental approach, the study by Gissurarson (1997) is a major landmark. He has asked participants in a series of PK experiments to describe after the PK task what technique, method or volitional strategy they used (I stay with the latter term in the rest of this paper). The very diverse descriptions he had obtained, were duly categorized and correlated with the scores on the PK task. The most successful strategy has indeed been one describeable as "resonance". His inventory of volitional strategies was the source for the choice of strategies in the present experiment.

The experiment was carried out as Phase 2 of the Giessen part of the multicenter PEAR/IGPP Mind Machine Interaction (MMI) replication study. The MMI replication study has been reported extensively with regard to Phase 1, in which 250 experimental sessions were carried out in each of the three labs, in Princeton, Freiburg and Giessen (Jahn et al., 2000). The overall results of this study have been, first, the expected scoring level falling short of significance, and second, definite indications of scoring patterns in the data, which unfortunately had not been predicted. Both in Freiburg and Giessen a second phase of similar size has been conducted. The reason to use the term "mind machine interaction" has been that when the participants may choose the target of a run, a positive outcome could be produced by precognition as well as by PK. However, the theoretical - not the psychological - distinction between these processes may be a moot point (Houtkooper, 2002). Nevertheless, the experiment was presented to the participants primarily as a PK task.

The present study was part of Phase 2 of the MMI replication in Giessen, as in this phase an exploratory study of volitional strategies was included. Scoring patterns were observed in the Giessen data, as has been reported earlier (Houtkooper, 2000). To give two examples of findings for the combined phases in Giessen, PK scoring increased significantly from participants' first session to the second ($z = 2.322$, $p = .02$, two-tailed) and secondly, within sessions a U-shape, or salience, scoring pattern was observed ($z = 2.349$, $p < .02$, two-tailed). Although these post hoc findings may be not too impressive with regard to the effect sizes involved, they were considered as indications for the presence of some PK effect in part of the data, enough to make the study of volitional strategies worthwhile.

METHODS

Software and hardware

The experiment was carried out using the computer software of the multi-center PEAR/IGPP study (Jahn et al., 2000), the PortREG computer program, run on a PC. The RNG used was the PEAR benchmark electronic Random Event Generator. This generator resulted in one 200-bit sample, or trial, being registered in a little less than one second. A short run, consisting of 100 trials, lasts about 85 seconds, while a long, 1000-trial run lasts almost 15 minutes. The PortREG program enabled the choice between sessions of 30 short runs or 3 long runs. Runs were balanced among the three different intentions, namely "High", "Low", or "Baseline" (in the following called H-, L- and B-runs, respectively). Furthermore the participant had the option of chosing the intention, or target direction, of each run him- or herself, or having the choice of intention made by the computer, using a pseudo-random algorithm, with a starting point derived from the computer clock, so that the sequence of intentions was different in each session. The pseudo-random algorithm produced a "well-shuffled closed deck" of the 30 intentions. Another option was the feedback mode of the random process. The standard condition of graphic feedback resulted in a graph depicting the cumulative deviation from chance expectancy from the beginning of the run, at the left edge of the screen, to the right edge of the screen at the end. The vertical deviation was proportional to the deviation from chance, so that a trial which consisted of 100 0's and 100 1's produced a horizontal stroke; if there were more than 100 1's the slope was upward, with more than 100 0's the slope was downward. This cumulative graph was colored red (on a black background) if the assigned intention, or target direction, was "High", blue if the intention was "Low" and green if it was "Baseline".

Presentation

The first session by each participant (Pt) employed the standard condition of graphic feedback, short runs and choice of intention by the Pt. In later sessions, the Pt could choose another condition freely. (Except for the differences between the long and the short runs, we do not discriminate here between the different options of the Portreg program.) It was deemed important to leave the protocol rather flexible, as a study of protocol-rigidity has indicated (Schneider, 1999). The volitional strategies were recorded by means of a protocol.

Volitional strategies:

First of all, the Pt was given an instruction by the experimenter about the 'willing' or 'wishing' he or she was going to do, and that people in earlier experiments had given their subjective impressions. The Pt was asked to keep a record of his or her performance on a form which has columns for intention and volitional strategy, both to be filled out before the run started, and the mean score of the run, which was a number around 100, and an optional column for comments. The volitional strategies to choose from were given as:

A: Imagery
B: Relaxed
C: Confident
D: Resonance
E: Guessing
X: Other, please describe your own strategy.

The text of the instruction was also given to the Pt in writing, see Appendix 1. The above choice of strategies was made, based on the most successful strategies in Gissurarson's study, as will be discussed later.

Subjects

In Phase 2 of the MMI-replication the participants were recruited by advertisements in a local newspaper, for an ESP experiment. One session was conducted, consisting of filling out a questionnaire and a 40-trial, forced-choice ESP test. The questionnaire has been described earlier (Houtkooper, 2003). The 88 participants in this ESP experiment were invited to take part in the MMI experiment: 26 males and 48 females volunteered. Their ages ranged from 19-60, while the mean age was 29 years (SD=8.1) with a median of 26.4 years. Per session, which typically lasted about an hour, the participants were rewarded with 15 DM or about 8 Euros.

Data collection

Data were collected from January to October 1999. The number of sessions per Pt was voluntary, up to

5. The 74 participants did a minimum of one session, the mean number of sessions being 3.66, for a total of 271 sessions. Volitional strategies were recorded appropriately, with very little missing data (less than 1% of the runs). Some Pts recorded volitional strategies also for the baseline runs, but this had little meaning for this study.

Hypothesis

The one pre-specified hypothesis for the present study was that the runs in which the Pt chose the "resonance" strategy, produced the best PK performance. In order to avoid confusion with a salience effect, this strategy was not put at the beginning or the end of the list, but rather in the middle.

Analysis

Student's t-test has been used throughout. For the calculation of within session declines and salience curves, t-values were converted to z-scores by tranformation of one-tailed p-values to z-scores. When the t-value is associated with many degrees of freedom, t and z have similar values. A z-score for decline was calculated by a linear weighting scheme for the order of the H- and L-runs as these are numbered 1 to 10 in the session. Similarly, a parabolic function was used to obtain the z-score associated with the salience pattern of the scoring during the session. The formulas are given in Appendix 2.

With regard to the novel approach to volitional strategies in the present study, all analyses here have to be regarded as exploratory, which is also the reason no corrections for multiple analysis are presented.

RESULTS

Strategies and PK scores

To evaluate the overall scoring in the experiment, the experiment was divided in 100-trial sections, that is, the long runs were divided into 10 sections of the same size as the short runs. The comparison between the 2710 sections with target high and the 2710 with target low, the "mean shift" in the average number of 1-bits per trial, revealed a nonsignificant difference in the intended direction (t=0.708, df=5418, p=.24, one-tailed). The results for the different volitional strategies are given in Table 1:

Table 1: Volitional strategies and their associated PK scores

Strategy	Mean shift	t	d.f.	p
A: Imagery	-0.017	-0.495	1609	.69
B: Relaxed	+0.005	0.115	967	.45
C: Confidence	+0.012	0.234	852	.41
D: Resonance	+0.115	1.889	577	.030
E: Guessing	+0.006	0.118	810	.45
X: Own strategy	+0.070	1.090	554	.14

Therefore, the hypothesis that the resonance strategy would perform best, has been confirmed, at p = .03, one-tailed. Apart from the six strategies there were 39 runs out of the 5420 where Pts did not indicate a strategy: On these few runs, a negative mean shift of -0.477 occurred with t(37) = -1.906 (p = .97).

Exploration of volitional strategy and PK score

Several questions arise with this finding: Are there further differences within the data? Is the resonance strategy chosen by – maybe a few – Pts who show certain properties? Do Pts often switch between strategies or do they stay with one dominant strategy? These questions were reason for some further explorations in the data.

An important condition with respect to volitional strategy was the difference between sessions with long and short runs. With the long runs, the Pt had to stay with his or her chosen strategy for about 15 min, without pause, so that the possibility of becoming distracted had to be taken seriously. Therefore, the relationship between strategy and PK scoring was analyzed for long and short runs separately. This resulted in nonsignificant scoring in the long runs for all strategies, whereas the result for the short runs was significant for the resonance strategy ($t(417) = 2.199$, $p = .014$, one-tailed) and marginally so for Pts' own strategies ($t(364) = 1.555$, $p = .060$, one-tailed).

Prevalences of volitional strategies

The prevalences of the volitional strategies in the present experiment might be compared with the verbal reports of the Pts in Gissurarson's study. We compare our percentages of runs with a particular strategy with his percentages of sessions within a particular category. As mentioned, the strategies in the present experiment were derived from his categories, based on the scoring rates in his experiments, taking the most successful strategies. The strategy which was the second best in Gissurarson's study, namely "external assistance", or "summoning outside entities", occurred so seldomly (in 1% of the sessions) that this was left out of the present study. Table 2 shows the comparison of the prevalences of strategies between the present experiment and Gissurarson's study.

Table 2: Prevalences, or percentages of occurrence, of strategies in the present experiment and those reported by Gissurarson (1997).

Strategy	Present Study	Gissurarson
A: Imagery	31.2	30
B: Relaxed	19.0	10.5
C: Confidence	14.4	5.5
D: Resonance	11.1	2.5
E: Guess	13.8	7
X: Other	9.7	1.5
Concentration/will	-	22
Physical strategies	-	11
Talking	-	7.5
Negative emotion	-	1.5
External assistance	-	1
Total	100	100
No reported strategy	0.7	10

The comparison of prevalences of strategies has been affected by the differences in the setting. The comparison of the scoring rates is given in Table 3.

Prevalences of strategies may be considered to depend on different situational or personal variables: Of these, two quite obvious ones, namely the difference between H- and L-runs (the B-runs have only spurious indications of strategy) and gender, reveal no large differences.

Table 3: PK scoring using different strategies in the present experiment (mean shift) and those reported by Gissurarson (average number of hits).

Strategy	Gissurarson	Present Study
A: Imagery	-0.017	10.29
B: Relaxed	+0.005	10.35
C: Confidence	+0.012	10.95[1]
D: Resonance	+0.115	13.67
E: Guessing	+0.006	10.05
X: Own strategy	+0.070	
Concentration/will	-	10.42
Talking	-	9.52
Physical strategies	-	9.91
Negative emotion	-	[2]
External assistance	-	11.50

[1]) “affective methods”
[2]) included with C: affective methods

Other scoring parameters

As mentioned in the introduction, there was a salience pattern present in the data, in the present study to a significant degree if tested one-tailed: $z_U = 1.900$, $p = .029$. This means that scoring tended to be positive in the beginning and at the end of the 10 H- and 10 L-runs, and zero or negative in the middle of the sequence. The data were also evaluated for a decline. This resulted in $z_{decl} = 1.597$, $p = .055$, one-tailed.

The data of the short runs in which the resonance strategy was employed, revealed no salience or U-shape pattern ($z_U = 0.001$), but the decline was significant ($z_{decl} = 2.401$, $p = .008$, one-tailed). That is, scoring dropped significantly from the first to the last short run of the session.

Psychological variables

The prevalence of different strategies certainly differs among Pts. In order to explore whether there are any systematic differences, the psychological variables obtained from the questionnaires have been examined. These are: The five personality factors of the NEO-FFI, the attitudinal variables of psi-belief, religiosity, risk taking and luckiness, the experiences with ESP, PK and training in a mental discipline and the answers to questions about the abilities to visualize and to concentrate or being absorbed. To analyze the relationship, first, on run-by-run basis, the psychological variables were trichotomized in low, middle and high scoring groups and the runs by Pts in each group were analyzed for relative prevalence of strategies and their scoring rates. (Trichotomization is a better alternative to the ubiquitous median split.)

For each variable, the contingency coefficient phi (Hays, 1973) between the trichotomized psychological variable and the volitional strategies was calculated. This analysis resulted in rather strong associations with the following variables: Luckiness, Visualization, Belief in PK (from the Australian Sheep-Goat scale), Religiosity and Conscientiousness and Openness from the NEO-FFI (Costa & McCrae, 1992). The first question was whether these associations also result in meaningful scoring differences and therefore these variables are presented here, starting with the variable which had the strongest association with the choice of strategy (phi=0.26), namely luckiness.

Table 4: Prevalences of strategies in runs by participants who rated themselves lucky to different degrees and the associated scoring parameters.

	Luckiness					
	High		Middle		Low	
Percentage runs with:						
Resonance		10.5		11.1		5.0
Imagery		22.2		37.0		45.3
Own strategy		18,1		14.6		3.0
Scoring on strategy:						
	t	p	t	p	t	p
Resonance:						
Overall t	1.941	.027	2.222	.014	-0.919	.8
Decline z	0.656	.3	2.125	.017	1.649	.050
Salience z	1.755	.040	-0.768	.8	-1.087	.9
Imagery:						
Overall t	0.822	.21	-0.739	.8	-0.887	.8
Decline z	1.196	.9	-1.099	.9	0.021	.5
Salience z	1.812	.035	-0.019	.5	0.263	.4
Own:						
Overall t	1.976	.024	-0.092	.5	-1.546	.9
Decline z	0.223	.6	0.734	.23	-	-
Salience z	0.228	4	-0.057	.5	-	-

Luckiness: The Pts who considered themselves to be highly lucky, chose resonance as a strategy in 10.5% of the runs, the middle group did so in 11.1% of cases, whereas the Pts who were low on luckiness chose resonance in only 5.0% of their runs. Further results on the contrasting strategies and related scoring parameters are summarized in Table 4.

It can be seen in Table 4 that the highly lucky Pts not only chose more often resonance as a strategy, but they were also quite successful in these runs. This significant finding (p=.027) is accompanied by a significant salience pattern (p=.040). The runs by "middle lucky" Pts show also significant overall scoring (p=.014), while significantly (p=.017) declining. The picture is less clear for the the imagery and own strategies.

Another question about these relationships is whether these are due to a few Pts who at all choose the resonance strategy, or that the relationship between luckiness and the choice of resonance is more general in all Pts in our study. To examine this aspect, the average number of choices of resonance per session was calculated for each Pts and this "resonanceness" was correlated with their luckiness score. This was also done for the other strategies and the correlation coefficients are shown in Table 5.

Table 5: Pearson correlations between some psychological variables and the prevalences of volitional strategies.

Psychological Variable / Strategy:	LUCKY	VISUAL	ASG-PK	Relig	Consc.	Openness
Imagery	-0.336**	-0.079	-0.087	-0.067	0.069	0.011
Relaxed	-0.042	0.027	0.078	-0.109	-0.049	0.113
Confidence	0.217+	-0.041	0.249*	0.264*	0.262*	-0.055
Resonance	-0.061	0.136	-0.198+	-0.092	-0.261*	0.081
Guessing	0.101	-0.115	-0.056	0.067	0.067	-0.288*
Own	0.240*	0.089	0.037	-0.020	-0.091	0.118

+: p<.10, *: p<.05, **: p<.01

As can be seen in Table 5 there were quite a few significant associations between psychological variables and volitional strategies. The correlation just mentioned, between Luckiness and resonance strategy, was however negative, -0.061, whereas from Table 4 a positive relationship would have been expected. The reason for this discrepancy is that the choice of resonance was very unevenly distributed amongst Pts: 5 of the 74 Pts contributed 40% of all resonance runs, whereas 24 of the 74 never chose the resonance strategy at all. On the basis of runs, those who chose resonance very often weighed heavily in the result, but on the basis of Pts these runs played a much smaller role.

Table 6: Prevalences of strategies for participants with the ability to visualize to different degrees and the associated scoring parameters.

	Visualization					
	High		Middle		Low	
Percentage runs with:						
Resonance	13.5		6.6		6.1	
Imagery	37.2		22.7		33.4	
Own strategy	9.9		20.4		12.9	
Scoring on strategy:						
	t	p	t	p	t	p
Resonance:						
Overall t	1.477	.07	1.390	.09	0.983	.16
Decline z	1.876	.03	1.638	.05	0.563	.3
Salience z	0.106	.5	0.409	.3	-1.037	.9
Imagery:						
Overall t	0.168	.6	-1.201	.9	0.374	.4
Decline z	1.579	.9	-0.053	.5	-0.356	.6
Salience z	0.939	.17	1.112	.13	0.160	.4

Own:						
Overall t	2.508	.007	-0.536	.7	0.970	.17
Decline z	0.110	.5	1.394	.08	-0.673	.7
Salience z	0.612	.7	0.499	.3	0.885	.19

Visualization: The ability to visualize, to produce imagery, has often been associated with PK tasks. The visualization variable was derived from four items in the questionnaire and the Pts were divided in three groups of more or less equal size according to their ability to visualize. Visualization has been associated with the imagery strategy and it was interesting to see if this association would reveal itself in the data. It is remarkable that with regard to visualization, the imagery strategy shows less prevalence in the middle group than in both the high and low visualization groups. The Pts' own strategies showed the opposite behaviour, see Table 6. In this Table, the only remarkable fact is that the relatively few high-visualizers who chose their own strategy, were rather successful. There was a relatively high preference for the resonance strategy with the high-visualizers and they did rather well, but their performance declined (p=.03) during the session. The obvious nonlinear relationship between visualization and the imagery and "own" strategies, is the reason that these relationships do not reveal themselves in the correlations in Table 5.

Conscientiousness: This personality variable showed relationships with two strategies which hold over all Pts, a positive correlation with the confidence strategy, whereas the higher conscientiousness, the less often the resonance strategy was chosen. These relationships apparently also hold when the runs are analyzed, see Table 7:

Table 7: Prevalences of strategies for participants with different levels of Conscientiousness and the associated scoring parameters.

	Conscientiousness					
	High		Middle		Low	
Percentage runs with:						
Resonance		8.0		5.7		14.5
Confidence		16.2		11.7		9.8
Scoring on strategy:						
	t	p	t	p	t	p
Resonance:						
Overall t	3.425	.0004	-0.980	.8	0.987	.16
Decline z	1.280	.10	1.437	.08	1.209	.11
Salience z	1.469	.9	-0.526	.7	1.688	.05
Confidence:						
Overall t	0.639	.3	0.809	.21	-0.965	.8
Decline z	1.470	.07	-0.379	.6	0.751	.23
Salience z	0.012	.5	-0.748	.8	-0.015	.5

The Pts with a high Conscientiousness scored highly significantly (p = .0004) in runs where they chose the resonance strategy. This finding on a by-run basis may be compared with a correlation among the high-conscientious Pts between their "resonanceness" and their overall PK-scoring. As it turned out, this correlation could be evaluated as a combined correlation with the PK-scores in the 5 sessions. This correlation was indeed positive (per session estimated r = 0.183, p = .032, one-tailed). On the basis of this correlation between participants, however, the effect appeared much weaker than on the basis of runs.

The other psychological variables which appear to be related to volitional strategies are not reported here. Their patterns do not reveal any more outstanding features than the three psychological variables mentioned so far.

"Own" strategies: The choice of a strategy for every run made it difficult to evaluate prevalences of different strategies. From the reported strategies, some consisted of combinations of given strategies, some were types of imagery, but most fell into Gissurarson's remaining strategies, see Table 2. Gissurarson's "external assistance" or "summoning outside entities" did not occur among the "own strategies" in the present study. The strategies occurring often were physical strategies, and strategies employing verbal commands and statements of concentration or will. A number of descriptions involved some process-oriented imagery, rhythmic breathing and keeping the eyes closed or squinting.

General Discussion

The reason for the present study was the perennial question "How do they do it?" in successful PK experiments. Furthermore, Gissurarson's study provided a major piece of groundwork, to which he present study is the next step forward. The prospective indication of a volitional strategy had hitherto been lacking and the present study fills this lacuna. The present finding that the resonance strategy is the most successful and significant ($p = .030$, one-tailed) corroborates the introspective findings of some of the early experimenters and more recently the members of the PEAR group, and especially the more systematic study by Gissurarson.

However, it has to be realized that there are important differences between the studies. Different PK tasks are one, but another important factor is the way volitional strategies are introduced. On the one hand we have Gissurarson, who interviewed his participants afterwards about their techniques. In the present study, a choice of five specific strategies was presented, but without obligation to actually use these. Even a "fake" volitional strategy was included, namely guessing what the machine would do. At the other extreme, one might go about it systematically, as customary in ordinary psychology experiments, and assign different strategies to different subjects or different sessions. Even if subjects would be asked afterwards about their preferences, it is clear that such a design is much more rigid. A finding in other work carried out in Giessen has suggested that a rigid protocol tends to produce null-results (Schneider, 1999). Also, when presenting volitional strategies, the role of suggestion should not be underestimated. Putting the resonance strategy in the fourth place of the list was done purposefully, not to draw special attention to it. This might be varied in future studies. The question of what the Pts actually did when they chose a particular strategy, cannot be answered beyond the choice they made. The strategies were discussed briefly with the Pts. When they asked about the meaning of, for instance, the word resonance, the only answer was that this was what successful participants had reported in previous experiments. In future studies, some elaboration might be called for, as well as a method of recording the choice which precludes making the choice after the run. As Pts were naive about expectations with regard to strategies and the instruction emphasized that all given strategies were associated with success in scoring, but that there were individual differences, the influence of the latter point is assumed to be negligible, but cannot be excluded in principle. More elaborate recording of volitional method might easily become deleterious to the experimental setting.

One disadvantage of a non-rigid protocol is the uneven distribution of runs with a certain strategy over the participants. This caused a puzzling contrast between Tables 4 and 5. My opinion is that we have to accept these peculiarities, because the standard-psychological, or agricultural, model of research is less likely to produce anything of interest. Nevertheless, some research on volitional strategies has been done in this way and not without some interesting findings. For instance, Morris, Nanko and Phillips (1982) derived from popular psychic training programs the idea to test a "goal-oriented" against a "process-oriented" visual imagery strategy. It turned out that the goal-oriented strategy appeared to be the more effective, but rather for those with no prior experience with mental development training and rather in those who could choose their preferred strategy. Again, suggestion might have played a role in the way of presenting the strategies to the participants and this could also have been the reason for the findings by Levy (1979) who found higher PK scoring for a goal- than a process-oriented strategy, but only with feedback. Without feedback, the difference was significantly in the opposite direction.

The present results indicate that psychological variables interact with the choice of volitional strategy. Self-perceived luckiness, and the conscientiousness scale of the NEO-FFI are examples, and have been scrutinized here. Differences in the preferences for different volitional strategies occurred between the groups of low-, middle- and high-lucky participants.

Those scoring high on luckiness chose more often the resonance, or their own strategy than the low-scorers on luckiness, while the opposite was true of the imagery strategy. This was accompanied by differences in scoring: The highly lucky scored on runs in which the resonance strategy was employed significantly above chance ($p<.03$), together with a significant ($p=.04$) salience pattern. The same group also showed significant above chance scoring ($p<.03$) when employing their own strategy.

The interactions between the ability to visualize and the imagery strategy were of interest: Those who could very well visualize chose most often the imagery strategy (37% of runs), but they were closely followed by those who scored low on visualization (33%), whereas the middle visualization group chose imagery as their strategy in only 22% of runs. PK scoring showed no significant differences among these groups on the imagery strategy. In contrast, the high-visualizers chose their own strategy relatively seldom, in 10% of their runs, on which they scored significantly ($p<.01$) above chance.

The conscientiousness scale of the NEO-FFI shows a similar pattern: Those scoring high on conscientiousness chose the resonance strategy less often, for 8% of the runs, but they achieved very significant scoring on those: $p<.001$.

As an overall analysis, we may observe 12 scores significant at the .05 level in Tables 4, 6 and 7. Although the decline and salience scores showed fewer significant scores than overall scores, the differences were not significant. Therefore, both declines and salience patterns might play a role as the "fingerprints" of psi, but not prominently so. If we compare the number of significant scores on runs with resonance strategy with the other strategies in the same tables, resonance reveals 9 significant scores out of 27, whereas the other strategies muster 3 significant scores out of 40. A chi-square on the 2x2 table turns out to be 8.1, with 1 degree of freedom, which is significant at the .005 level. Therefore, we may conclude that the resonance strategy prominently shows scoring patterns in the interaction with some of the psychological variables.

What further interpretation might be given to the resonance strategy? Apart from the apparent absence of conscious striving, a probable lessening of ego-involvement, there is also the association with absorption. This association may be further scrutinized by looking at the one item in the questionnaire, originally labeled "concentration". The question involved was: "How often are you so absorbed in an activity that you forget your surroundings?" (The German word "vertieft" is best translated by "absorbed"). Answers were on a 7-point scale with "never" and "always" as end-points. When calculating the correlation with the prevalence of the resonance strategy, we left out the participants who chose resonance never or not more than once every two sessions, as being irrelevant for the relationship with absorption. The correlation between prevalence of resonance and absorption then turned out to be significant: $r = .293$, $N = 40$, $p < .04$, one-tailed. We may therefore surmise that there is an underlying relationship with the choice of the resonance strategy and the ability to be absorbed. This suggests a better performance under a condition of absorption, which would be consistent with findings in research on experimental and spontaneous ESP, but apparently not with what is known about spontaneous PK experiences (Irwin, 1985).

The "own" strategies were often suggestive of magical actions. One might be reminded of Beloff's history of parapsychology, where he gives one of the definitions of magic as: "the employment of ineffective techniques to allay anxiety when effective ones are not available" (Beloff, 1993). The question remains here whether the ineffective techniques may elicit an effective state of mind.

An unexpected result of the study is a question: "Are we looking for the generalizable properties of persons taking part in an experiment, or are we looking for the generalizability of the state of mind of different people when they say that they are using a resonance strategy?"

This has come forward as a result of the explorative analyses, of the interactions of volitional strategies with both luckiness and conscientiousness. I must say that from the beginning I had a model in mind in which PK performance is a property of the person, which may vary somewhat. Now I conclude that the momentary state of mind of the participant in these experiments may play the bigger role. Follow-up studies might attempt to distinguish between overall preference for a strategy and the momentary choice of a strategy by the participant.

ACKNOWLEDGMENTS

Financial support by the Institut für Grenzgebiete der Psychologie und Psychohygiene is gratefully acknowledged. Thanks are due to Simone Launspach, Dipl. Psych., and Stefan Schneider, Dipl. Psych., who took part as experimenters in the experiment and who contributed to the data analysis in a dedicated manner.

REFERENCES

Beloff, J. (1993). *Parapsychology: A concise history.* London: Athlone.

Costa, P. T., & McCrae, R. R. (1992). *The NEO personality inventory manual.* Odessa, FL: Psychological Assessment Resources.

Forwald, H. (1952). A Continuation of the Experiments in Placement PK. *Journal of Parapsychology, 16*, 273-283.

Forwald, H. (1969). *Mind, Matter, and Gravitation.* Parapsychological Monographs No.11. New York: Parapsychology Foundation.

Gissurarson, L. R. (1997). Descriptive analysis of mentations on volitional tasks. *Journal of the Society for Psychical Research, 62*, 22-35.

Hays, W. H. (1973). *Statistics for the social sciences.* London: Holt, Rhinehart & Winston.

Houtkooper, J. M. (2000). Mind-Machine Interaction: Two Phases of replication in Giessen. *Proceedings of the 43rd annual convention of the Parapsychological Association* (Freiburg, August 17-20, 2000), 419.

Houtkooper, J.M. (2002). Arguing for an observational theory of paranormal phenomena. *Journal of Scientific Exploration, 16*, 171-185.

Houtkooper, J. M. (2003). A forced-choice ESP experiment with natural and simulated sferics: Displacement scores and psychological variables. *Proceedings of the forty-sixth annual convention of the Parapsychological Association* (Vancouver, Canada, August 2-5, 2003), 82-97.

Irwin, H. J. (1985). Parapsychological phenomena and the absorption domain. *Journal of the American Society for Psychical Research, 79*, 1-11.

Irwin, H. J. (1994). *An introduction to parapsychology* (2nd edition). Jefferson, N.C.: McFarland.

Jahn, R. G., & Dunne, B. J. (1987). Margins of reality. The role of consciousness in the physical world. San Diego: Harcourt Brace.

Jahn, R., Dunne, B., Dobyns, Y., Lettieri, A., Nelson, R., Mischo, J., Boller, E., Bösch, H., Vaitl, D., Houtkooper, J. M., & Walter, B. (2000). Mind/Machine Interaction Consortium: PortREG Replication Experiments. *Journal of Scientific Exploration, 14*, 499-555.

Levi, A. (1979). The influence of imagery and feedback on PK effects. *Journal of Parapsychology, 43*, 275-289.

Morris, R. L., Nanko, M., & Phillips, D. (1982). A comparison of two popularly advocated visual imagery strategies in a psychokinesis task. *Journal of Parapsychology, 46*, 1-16.

Schmidt, H. (1997). Random generators and living systems as targets in retro-PK experiments. *Journal of the American Society for Psychical Research*, 91, 1-13.

Schneider, S. (1999). Einfluss psychologischer Variablen und des experimentellen Settings auf die Scores eines Psychokinese- und Präkognitionstests. [The influence of psychological variables and the experimental setting on the scores on a psychokinesis and a precognition test.] Master's thesis, Department of Psychology, University of Giessen.

Stanford, R. G. (1977). Experimental psychokinesis: A review from diverse perspectives. In: B.B. Wolman (Ed.) *Handbook of parapsychology.* New York: Van Nostrand Reinhold. (pp. 324-381)

Thouless, R. H. (1951). A Report on an Experiment on Psycho-Kinesis with Dice, and a Discussion of Pschological Factors Favouring Success. *Proceedings of the Society for Psychical Research, 49*, 107-130.

Address for correspondence: Joop H. Houtkooper, Center for Psychobiology and Behavioral Medicine, Justus Liebig University of Giessen, Otto-Behaghel-Strasse 10, 35394 Giessen, Germany.
E-mail: joop.m.houtkooper@psychol.uni-giessen.de

APPENDIX 1

Instruction about volitional strategies
(translated from the German)

Persons who had good results in psychokinesis experiments, often described their “method” as:

A. Imagining a picture of a positive result.
B. Being attentive and at the same time relaxed.
C. A positive feeling or confidence about the wanted result.
D. A feeling of resonance with the random generator and/or the computer.
E. Guessing/predicting, in which direction the result will be.

Generally you can decide each time in the experiment, which method best decribes what you will be doing. You may choose one of the above mentioned options, or you may choose your own method.

Before you start a run, note your intention and your method. Thereby you can use the above letters A to E, respectively X, if you are going to use your own method. Please describe your own method on the reverse side of the protocol. If you want to refrain from noting your method, just put a stroke in the column “Method”.

Note: During a run, stay with the method you have chosen, also when it seems that it doesn’t bring the wanted result.

After each run, please write down the mean score and if you want you may comment on your experience during the run.

Note: In the Tables, the following abbreviations are used:
A: Visual; B: Relaxed; C: Confidence; D: Resonance; E: Guessing; X: Other.

APPENDIX 2

Z-scores for decline and U-shape scoring pattern.

A session in the experiment consists of 10 H-, 10 L- and 10 B-runs. The difference between H- and L-runs is tested by the t-test, testing the first H-runs against the first L-runs and so on. The one-tailed p-values associated with the t's are transformed to *z*-scores. This results in 10 *z*-scores. From these, the *z*-score for decline may be calculated as:

$$Z_{decl} = (9z_1 + 7z_2 + 5z_3 + 3z_4 + z_5 - z_6 - 3z_7 - 5z_8 - 7z_9 - 9z_{10})/\sqrt{330}$$

The U-shape or salience pattern tests for a quadratic function (parabola) as the dependency of scoring on position of the run within the session. The *z*-score for salience pattern is:

$$z_U = (14z_1 + 5z_2 - 2z_3 - 7z_4 - 10z_5 - 10z_6 - 7z_7 - 2z_8 + 5z_9 - 14z_{10})/\sqrt{748}$$

Thus, the overall scoring, decline and salience pattern are three independent components of the changing scoring with position within the session.

DISTANT PSYCHOPHYSIOLOGICAL INTERACTION EFFECTS BETWEEN RELATED AND UNRELATED PARTICIPANTS

Marios Kittenis[1], Peter G. Caryl[2], & Paul Stevens[1]
[1]*Koestler Parapsychology Unit, University of Edinburgh*
[2]*Psychology Department, University of Edinburgh*

ABSTRACT

The aim of this study is to investigate possible remote psychophysiological interactions between sensorially isolated participants, using a protocol of photic stimulation and EEG measurements. It is an attempt to conceptually replicate past findings suggesting the presence of such interactions, and to clarify the role, (if any), of an existing emotional relationship and pre-session interaction between participant pairs.

Forty-one unpaid volunteers were assigned to one of three groups. One of these consisted of thirteen related pairs of participants who reported sharing an empathic relationship, another of five unrelated pairs (i.e. randomly matched strangers), and the last of five single participants. Related pairs spent some time alone together before testing, whereas unrelated pairs did not know each other and did not meet until after the session; single participants were told they would be paired with someone they didn't know, but were not matched with anyone. Pairs of participants simultaneously listened to a recording of a progressive relaxation procedure including suggestions aimed to induce a hypnagogic-like state, which was followed by 15 minutes of continuous drumming; this procedure was intended to induce a similar alteration of consciousness in both participants. During the drumming period the EEG of one person of the pair ("receiver") was recorded while the other ("sender") was occasionally stimulated with randomly timed single photic flashes. For the single participants group the same procedure was followed but there was no "sender" to observe the flashes.

EEG epochs that were time-locked on photic stimulation of the "senders" were taken from the continuous EEG record of the "receivers". Similar randomly sampled epochs were taken from periods of no stimulation to serve as controls. According to the null hypothesis no difference would be expected between these samples, as sensory stimulation of the "receivers" was homogenous throughout the experimental period. Event-related evoked alpha power measures revealed a tendency for samples from "remote" photic stimulation periods to show larger deviations from pre-stimulus baseline than control samples; these deviations were in the same direction as normal responses to direct photic stimulation. This difference between "remote" photic stimulation and control periods was found to be significant for the related pairs group at $p<0.023$ (Wilcoxon signed-ranks test, two-tailed; $N=13$). Deviations of similar direction and magnitude were found in unrelated pairs ($p<0.007$ when combined with related group, $N=18$), while recordings from single participants (when no other person was stimulated) showed no such effect. Further patterns identified in the results and possible interpretations are discussed.

INTRODUCTION

An increasing number of parapsychological experiments make use of physiological measures as dependent variables, rather than the more traditional psychological/behavioural measures. This choice relies on the fairly reasonable assumption that perception of any psi-mediated information will inevitably result in measurable changes in physiological parameters at some stage of the perceptual process, as is the case with ordinary perception. An additional motive for using such measures is the possibility that directly measuring physiological parameters may be advantageous in detecting subtle non-normal perceptual responses, which are perhaps not salient enough to rise above the threshold of conscious awareness. For psi experiments using psychophysiological recordings that directly measure parameters of brain activity, such as the EEG and fMRI, an extra potential advantage is that such measures could possibly help in identifying some of the underlying neural mechanisms that may be involved in psi processes. Of particular relevance to the present study is research using visual stimulation of one participant while measuring the EEG of another non-stimulated subject.

In one of the first studies to use such methodology, "senders" were stimulated with intermittent photic flashes at 6 and 16Hz. Seven pairs were tested, and while in only one of these the "receiver" showed alpha power blocking when the "sender" was stimulated, this pair was tested further and a repeatable effect was observed (Targ & Puthoff, 1974). In a subsequent study, pairs of subjects meditated together and were then taken to separate Faraday cages. One person of each pair was stimulated with trains of 100 flashes at random time intervals, while EEG was recorded from both. The stimulated subjects demonstrated visual evoked potentials as expected, which significantly correlated with the EEG activity of the non-stimulated subjects, which were said to demonstrate "transferred potentials" (Grinberg-Zylberbaum, Delaflor et al., 1994). Apparently this was not the case for the control condition, in which the subjects in each pair did not interact prior to the experiment. A subsequent attempt at a replication of this study by Sabell, Clarke et al., (2001) however, failed to find an effect. A more recent study using a similar experimental design but different stimuli and analysis methods, has reported significant deviations from baseline in non-stimulated participants when another was visually stimulated, but found no differences in effects between related and unrelated pairs and nothing resembling a "transferred potential" (Wackermann, Seiter et al., 2003). A conceptual replication by Radin, (2003) further suggested that "receivers" were indeed responding to the "senders" being stimulated and not to the distant stimuli themselves. Such interactions can perhaps more correctly be described as correlations rather than "remote responses", especially as it is not yet clear whether they exhibit the same physiological characteristics as responses to direct stimulation.

This project is intended to be a conceptual replication of these previous studies, an attempt to further explore the nature of these effects, and an attempt to clarify the issue of whether interpersonal relationship and prior interaction between participants is a variable affecting the observed effect.

METHOD

Design:

EEG was recorded from one participant while another spatially isolated subject was stimulated with photic flashes. Our participant pool consisted of three groups, involving empathically related pairs, unrelated pairs, and single subjects. Randomly timed stimuli (single photic flashes) were presented interspersed with randomly timed control events, and event-related band power measures were used to compare stimulation and control epochs. The null hypothesis predicts no differences between such epochs for the unstimulated person of the pair, and no differences between groups. Individual sessions, where each participant was directly stimulated while their own EEG was recorded, were also conducted in order to investigate the normal physiological responses to these stimuli.

Participants:

Forty-one unpaid volunteers took part in the study, divided into three groups; thirteen related pairs i.e. pairs of volunteers who reported sharing an empathic relationship (close friends, relatives or lovers), five unrelated pairs (i.e. ten individual volunteers who didn't know each other were randomly matched into pairs), and five single subjects (individual volunteers were not matched with another, although they were told they would be, i.e. there was no "sender"). There were 23 female and 18 male participants, with a mean age of 28.7, ranging between 20 to 58 years of age.

Audio materials:

We have attempted to alter the conscious state of our participants using an audio recording of a progressive relaxation procedure, which included suggestions for entering a hypnagogic-like state, followed by a recording of continuous drumming with an inter-beat frequency ranging between 4-5Hz (recording of live drumming). The aim of this procedure was to facilitate deep relaxation, to induce a non-ordinary conscious state simultaneously in both participants, and to help participants maintain an awareness of each other, while also giving them suggestions for avoiding any effort to succeed in the task and for suspending any positive or negative expectations they may have.

System implementation:

A 40 channel NuAmps EEG system (Neuroscan, USA), was used for data acquisition and analysis. Thirty-two monopolar EEG channels were recorded (including references) from the following electrode sites: Fp1, Fp2, Fz, F3, F4, F7, F8, FCz, FC3, FC4, FT7, FT8, Cz, C3, C4, T7, T8, CPz, CP3, CP4, TP7, TP8, Pz, P3, P4, P7, P8, Oz, O1, O2 , A1 and A2. The reference was linked ears and we sampled at a frequency of 500Hz. A 50Hz bandstop filter was used, and the bandpass filter range was 1-100Hz. An electrode cap was used for electrode placement together with clip ear electrodes; all electrodes were sintered Ag/AgCl.

To present photic stimuli we used a pair of dark glasses fitted with eight white (clear) LEDs, (four over each eye). Photic flashes were triggered using TTL pulses from the parallel port of a PC running a script-driven program (Inquisit by Millisecond Software), which controlled the randomised presentation of two types of stimuli, one of which would trigger a flash and simultaneously register an event marker on the EEG trace, while the other (control event) would only set an event marker on the EEG trace with no associated flash presented. Inquisit used a pseudo-random algorithm to sample with replacement one of the two stimuli, and one of eight possible inter-stimulus delays of 1-8 seconds (i.e. mean IS interval was 4.5s). One hundred and eighty-six stimuli were presented during each joint session; on average half of these (93) would be single photic flashes and half would be control event markers on the EEG trace. Individual sessions (direct photic stimulation) consisted of 68 stimuli of each type, with the same range of randomly chosen inter-stimulus intervals.

The audio recording was played to both participants using a shared one-way audio link. The computer controlling stimulus randomisation was connected to the "sender's" LED glasses and to the EEG amplifier in the "receiver's" room and used synchronised TTL pulses to trigger flashes and set event markers on the EEG record, marking the timing of photic flashes and control periods. TTL inputs to the EEG amplifier are isolated from the participant and the amplifier, protecting against contamination of the EEG record from the electrical signals used to provide event markers, and of cueing the participants to the existence or timing of these signals. No auditory or visual cues were emitted from the amplifier that could indicate the presence of the triggers to the "receivers".

Procedure:

In individual sessions each participant was directly photically stimulated while his or her own EEG was recorded. In "remote" sessions, the EEG of one participant was recorded while the other (or no one in the case of the "no partner" group) was photically stimulated.

Related pairs of participants decided themselves who was to be the "sender" and who the "receiver", either by choice or randomly. They could spent 10-15 minutes alone together before the session, doing anything they thought might help them enhance their awareness of each other. Some possibilities were suggested, such as joint meditation, synchronised breathing, exchanging personal items (e.g. jewellery), but were encouraged to do whatever felt most appropriate for them both. They were discouraged from using verbal interaction during this period, and they were given the option to burn some incense while in the room together, of which they could each take some in their respective separate experimental rooms. This was thought to be likely to help them maintain any feeling of "connectedness" they may have achieved in the common room into the experimental period, as odours are especially effective as memory cues and are particularly effective in evoking the emotional elements of memories. Evidence from fMRI studies for example, indicates that the subjective experience of the emotional potency of odour-evoked memories is correlated with specific activation in the amygdala, which is greater in magnitude than that seen when the same memories are evoked using visual cues (Herz, Eliassen et al., 2004). A common odour in participants' respective rooms would also make their sensory environments more similar. Participants were given a choice between different types of incense, and most pairs (but not all) opted to use some. After spending time alone together, participants went to their respective experimental rooms and did not interact with the experimenter again, (or anyone else), until the end of the session.

Unrelated pairs did not know each other prior to the experiment and only met after the session had finished. Therefore the experimenter chose randomly who was to be the "sender" and who the "receiver".

The five single participants who were not matched with a "sender" were told that they would be paired with someone they didn't know, and that they would meet them after the experiment (i.e. the same as what the unrelated pairs were told). They were all "receivers", and while the photic stimulation procedure was carried out as described above, there was no "sender" in the other room to observe the flashes. After the session the experimenter gave these participants a full debrief and explained the reasons for the deception.

At the beginning of each session the progressive relaxation instructions were played to the participants; this recording lasted for approximately 11 minutes and was followed by the drumming, which lasted for approximately 15 minutes. Two minutes after the drumming had started randomised photic stimulation was initiated, which lasted for an average of 11.7 minutes (the actual session length depending on the cumulative duration of the randomly chosen inter-stimulus intervals).

RESULTS

The raw EEG data from all N = 41 participants was treated with a 1-30Hz band-pass filter and visually inspected for artefacts. Channels that were consistently noisy or which lost electrode contact during recording were marked and excluded from further analysis. The entire EEG records of two participants had to be excluded from analysis due to faulty recording (loose reference electrode). One of these was a "sender" during direct photic stimulation, and the other a "receiver" (from the "alone" group), during remote photic stimulation.

Three-second long epochs were taken from the continuous EEG records, centred on stimulus presentations times (and random control markers) ranging from -1 to +2 seconds. According to the stimulus randomisation protocol we had used, the shortest possible interstimulus interval was 1s; therefore we could not use all of the 3s epochs, as some would contain more than one stimulus event and/or overlapping responses to stimuli. We therefore excluded from our analysis all events appearing after inter-stimulus intervals of <3s. Epochs were baseline corrected and those containing amplitudes >100μV were automatically rejected; epochs were also visually inspected and those found containing additional artefacts from eye movements or muscle activity were manually rejected. This manual artefact rejection was conducted blind as to whether epochs contained photic or control events. After such rejections, the number of epochs of each type available for analysis for each person and channel averaged at 55 for direct photic stimulation sessions and 70 for "remote" stimulation sessions. (The average number of stimulus events originally presented was 67 and 93 respectively).

The EEG data from direct photic stimulation sessions was analysed first in order to investigate the electrophysiological characteristics of normal responses to the photic stimuli we were using, and thus provide a template with which to guide the analysis of data from "remote" sessions.

Results of direct photic stimulation sessions:

Event-Related Band Power measures (ERBP) were used, where the raw EEG of all event-related epochs is band-pass filtered around a central frequency band of interest. We have chosen to focus on the alpha band (8-13Hz), as power in this band is well known to be affected by photic stimulation, e.g.(Kawaguchi, Jijiwa et al., 1993). We have used a measure of evoked (phase-locked) activity, as initial analysis demonstrated that evoked responses to photic stimulation are better defined and simpler to describe than induced (non phase-locked) responses. In evoked ERBP, the amplitude values within each epoch are squared in order to obtain power measures (μV^2), and a number of epochs that are time-locked to the stimulus are averaged point-by-point (as in Event Related Potentials). Evoked alpha ERBP measures can therefore effectively be described as the specific alpha-band component of the general ERP.

The Global Field Power (GFP) was calculated for each participant from the 30 original electrode channels as a measure of global EEG activity. The GFP is defined as the standard deviation across multiple channels as a function of time, and is used to quantify the instantaneous global activity across the spatial potential field sampled over the scalp (Lehmann & Skrandies, 1980). An example of the GFP of the evoked alpha response to photic stimulation can be seen in Fig.1, showing a rapid increase in alpha power which starts almost immediately after stimulus presentation (*T*=0ms), peaks at 224ms and returns to baseline near 500ms after stimulus presentation.

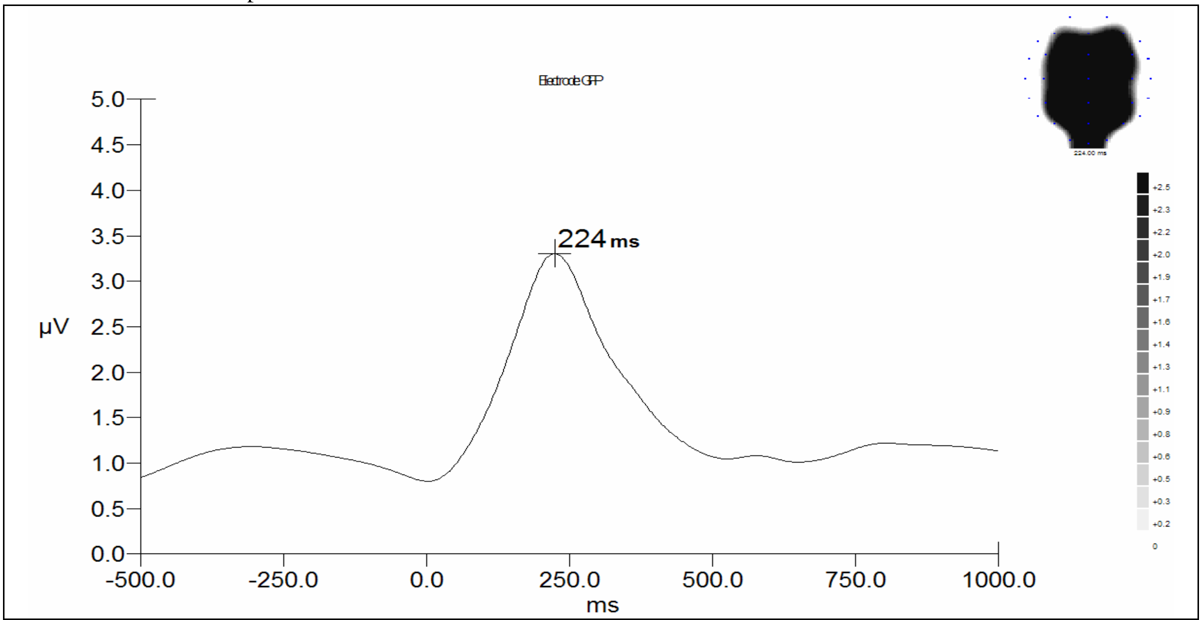


Figure 1: Evoked alpha response to direct photic stimulation; averaged Global Field Power for 30 channels and N=39 Ss.

We can define a period of interest within remote stimulation epochs based on the averaged responses of all participants to direct photic stimulation, on the assumption that if "receivers" are responding to photic stimulation of the "senders", their responses will have similar temporal characteristics. We therefore defined our test interval to be the range of 0 to 500ms after stimulus presentation, as responses to direct stimulation reach a maximum and return to baseline within this interval. As a comparison reference period we used the pre-stimulus interval of –500 to 0ms. We can calculate a ratio measure of post- to pre-stimulus power using the formula: $10^* \log \frac{W_{post}}{W_{pre}}$

where W_{post} is the mean α-power in the 0 to 500ms post-stimulus interval and W_{pre} is the mean α-power in the –500 to 0ms pre-stimulus interval. Therefore if there is no difference between pre-stimulus and post-stimulus power the log-ratio value would be 0, whereas positive values would indicate a higher mean α-power in the post-stimulus interval, and negative values would indicate a higher mean α-power in the pre-stimulus interval. For example, the log-ratio of such a comparison for the response to direct photic stimulation seen in Fig.1 would be: 10*log (1.90 / 1.03) = 2.6.

Results of "remote" photic stimulation sessions

As we would expect no systematic difference in α-power between pre- and post-stimulus intervals in the EEG of the "receivers", for the simple reason that they are not being stimulated themselves, we could in theory simply compare the above log-ratio of evoked α-power measures from epochs time-locked on photic stimulation of the "senders", against the expected value under the null hypothesis, i.e. 0. As we do not know however the exact statistical properties of the EEG signal, such theoretical assumption may not be justified, and a safer route would be to compare the log-ratio from periods of photic stimulation of the "senders", against the same ratio from control periods of no stimulation. The null hypothesis would also predict no difference between such periods for the unstimulated "receivers". Figure 2 shows the mean estimated log-ratios of these intervals for the three groups and two conditions.

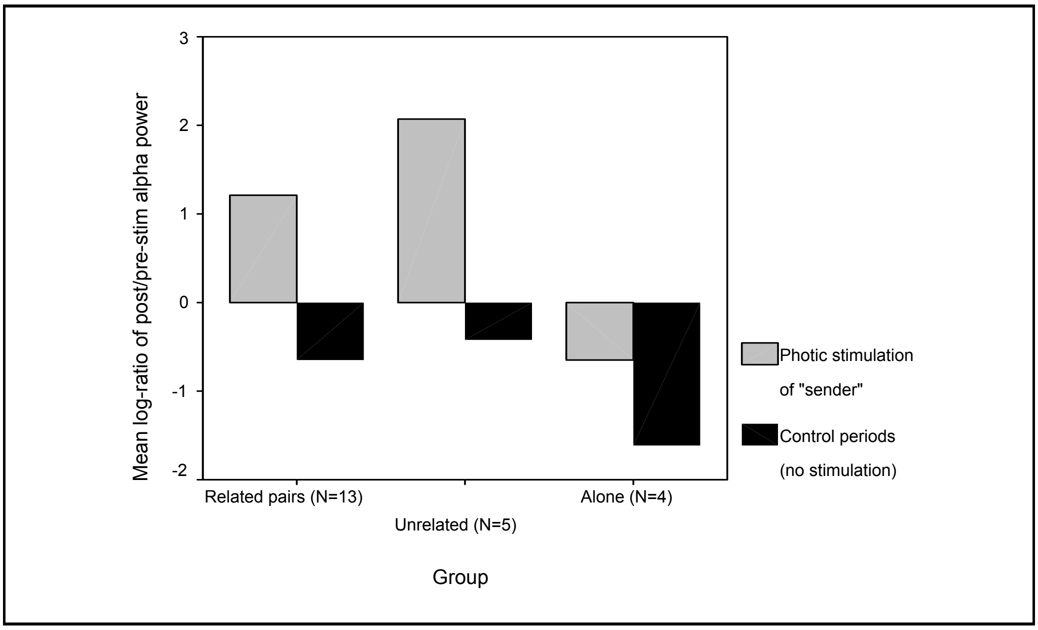


Figure 2: Mean log-ratios of post/pre-stimulus α-power (GFP) per group and condition.

A trend can easily be identified for positive ratios to be observed during epochs of photic stimulation of these participants' partners, for both related and unrelated pairs, indicating higher alpha power in post-stimulus intervals compared to the pre-stimulus baseline. No such difference can be seen in participants who were not matched with a partner (no one seeing the flashes); the log-ratios in this group are negative in both photic and control periods, indicating higher alpha power in the pre-stimulus interval. This is also the case for the control periods in the related and unrelated groups. Somewhat unexpectedly the deviation from baseline in photic periods appears larger for unrelated pairs, but the small number of participants in this and the "alone" group requires that we make such comparisons between groups only tentatively and with much caution.

These differences are highly comparable with recent findings in other studies, which have identified a similar pattern of effects between groups similar to these. Wackermann, Seiter et al., (2003) found deviations from baseline activity in the EEG of non-stimulated subjects, coinciding with periods when their partner was visually stimulated. Groups of related and unrelated pairs showed responses of similar magnitude, while a group of participants having no partner, and another with pairs where the "sender" could not see the stimuli, did not show any such responses. It is important to note that as in that study different visual stimuli (pattern reversal) and different EEG analysis methods were used to ours, the agreement between them is therefore only of a qualitative nature.

Table 1: Overall mean log-ratios of post/pre-stimulus α-power and standard deviations for each of the three groups and two conditions.

	Group	Mean	Std. Deviation	N
"Remote" photic stimulation periods	Related	1.22	1.9	13
	Unrelated	2.07	2.17	5
	Alone	-.65	1.33	4
	Total	1.07	2.01	22
Control periods (no stimulation)	Related	-.64	1.96	13
	Unrelated	-.41	1.76	5
	Alone	-1.62	3.07	4
	Total	-.77	2.08	22

To test the statistical significance of the observed difference we used a Wilcoxon matched-pairs signed-ranks test, which is distribution-free and does not rely on parametric assumptions. As can be seen in table 2 below, the difference in the related pairs group between photic and control epochs is significant at $p<0.023$ ($N = 13$; two-tailed). Conducting the test on groups as small as the unrelated pairs ($N = 5$) and unpaired participants ($N = 4$) is not likely to be reliable and will not be attempted. We could however combine the results from related and unrelated pairs, in which case we find a $p<0.007$ ($N = 18$; two-tailed). The overall difference between photic and control epochs for all three groups combined is also significant at $p<0.007$ ($N = 22$; two-tailed).

To estimate the effect sizes associated with these differences we calculated the values of the effect-size correlation r using the following formula:

$$r = d\Big/\sqrt{d^2 + 4} \qquad \text{where} \qquad d = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(\sigma_1^{\,2} + \sigma_2^{\,2})/2}}$$

Table 2 shows the calculated effect sizes and p values for each group and combination of groups.

Effect sizes of $r > 0.3$ are considered to be large; for example the $r = 0.43$ seen in the related group indicates that the mean of the photic condition stands at the 84th percentile of the control condition. Such effect sizes are comparable to some of the largest found in DMILS studies, where for example, the average effect size for 19 such experiments was found to be $r = 0.25$ (Schlitz & Braud, 1997).

Table 2: Estimated effect sizes and p values for differences in evoked alpha power changes between control and photic conditions; calculated for all groups separately and in combinations.

	Related Pairs (N=13)	**Unrelated Pairs** (N=5)	**"No sender" group** (N=4)	**Related & Unrelated Pairs** (N=18)	**All three groups combined** (N=22)
Wilcoxon Signed - Ranks Test	$p < 0.023$ (2-tailed)	N/A	N/A	$p < 0.007$ (2-tailed)	$p < 0.007$ (2-tailed)
Effect size *r*	$r = .43$	$r = .55$	$r = -.1$	$r = .47$	$r = .41$

It will be useful at this point to look at the temporal and spatial characteristics of the averaged waveform of alpha-ERBP from the "remote" stimulation periods. Figure 3 shows the average Global Field Power for the two groups that appeared to show an effect, i.e. the related and unrelated pairs (*N*=18).

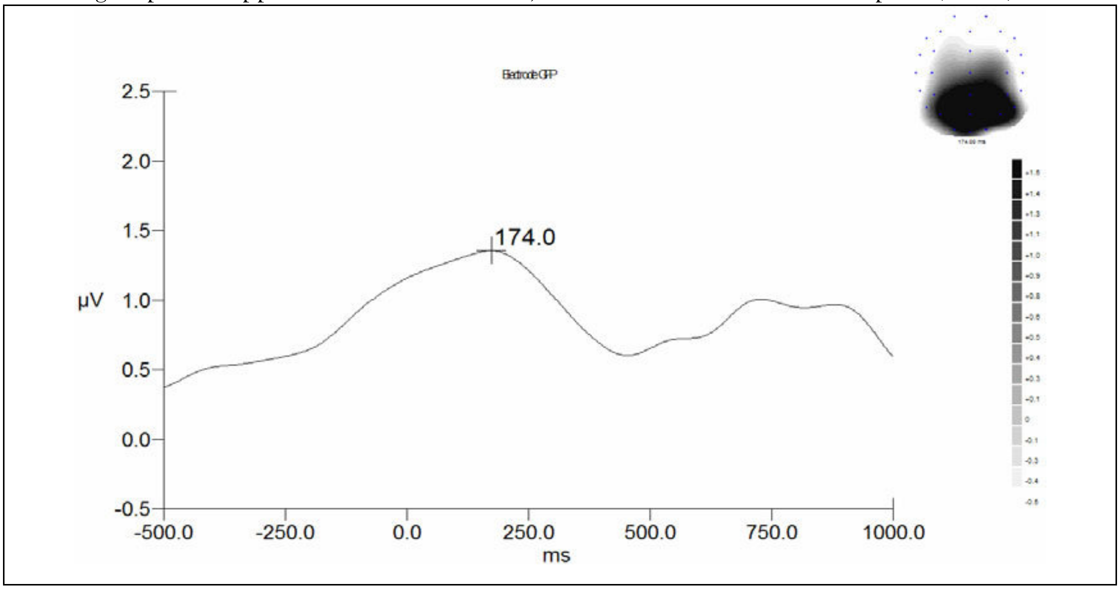


Figure 3: Mean Global Field Power of alpha-ERBP during "remote" photic stimulation for related and unrelated "receivers" combined (N=18).

This shows a relatively slow rise in phase-locked alpha power which peaks at 174ms post-stimulus; this "response" however also appears to begin to rise at least -150ms before stimulus presentation. Such a feature is obviously highly problematic if we attempt to interpret the effect as a physiological response to a remote stimulus, as it would violate the generally accepted assumption of linear causality, according to which responses follow stimuli and not vice versa. The spatial distribution of the effect however (see Fig.4) indicates a parietal/occipital locus for the observed deviation, which would be expected for the alpha component of a visual evoked response. Therefore unlike its problematic temporal characteristics, the posterior localisation of the effect is somewhat consistent with a physiological interpretation. The temporal evolution of the "remote response" can also be seen in Fig.4 and can be compared to that of the normal response to direct stimulation.

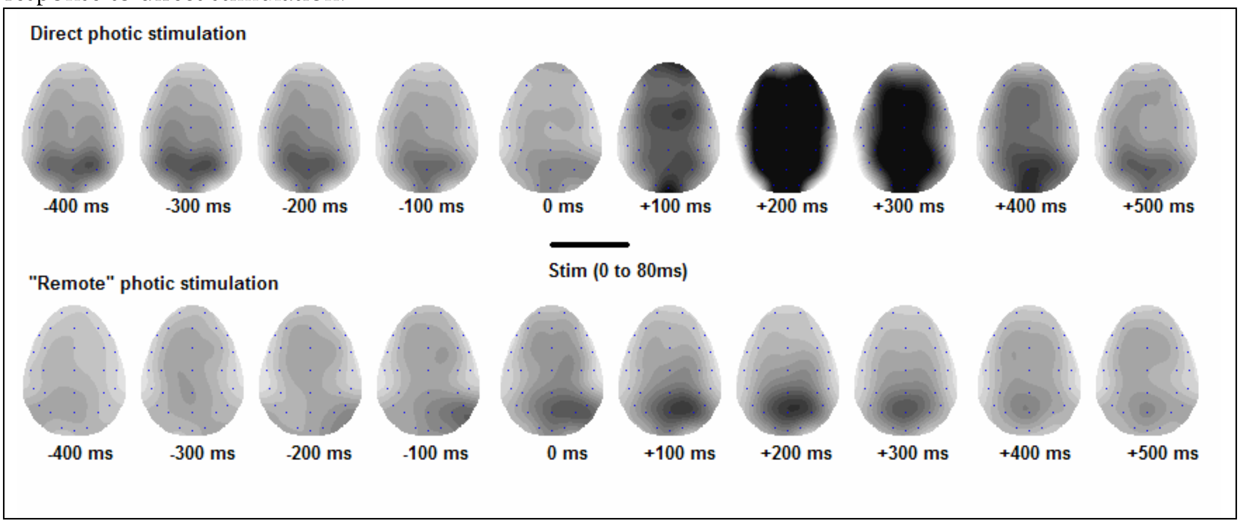


Figure 4: Spatial distribution of evoked alpha-power during direct photic stimulation (N=39) and during "remote" photic stimulation (related & unrelated pairs, N=18). Photic flashes were presented at T=0 for 80ms.

We are now in the process of analysing the results from the questionnaires administered to our participants, i.e. the *Modified Tellegen Absorption Scale* (Jamieson, 1986), the *Phenomenology of Consciousness Inventory* (Pekala, 1991), and a general participant information form, to further explore individual variables that may be related to performance in this experiment. Particular attention will be paid to the reported subjective consciousness alterations (in the PCI), and whether their intensity and quality, as well as correlations between the experiences of participants in each pair, relates in any way to task performance.

DISCUSSION

The results indicate that at times when the "senders" are stimulated with photic flashes the EEG of the "receivers" shows changes in α-power in the same direction as that seen when participants are themselves stimulated with photic flashes, i.e. phase-locked α-power increases during the post-stimulus period. As sensory stimulation for the "receivers" is homogenous throughout the experimental period, the significant difference between "remote" photic stimulation epochs and control epochs in log-ratios of post-/pre-stimulus α-power, suggests the presence of an anomalous effect during the "remote" photic stimulation periods. The lack of such a difference in the group of "receivers" who were not paired with a "sender", further suggests that this effect is dependent on sensory stimulation of another participant, and cannot be attributed to a general methodological flaw, or to direct extra-sensory perception of the remote stimuli. The parietal/occipital locus of the effect is consistent with what would be expected from the alpha component of a visual evoked response, and further suggests the presence of a "remote response".

The time-evolution of the observed effect however, is somewhat problematic and raises additional questions. The fact that "remote responses" appear to start at around -150ms pre-stimulus, and to peak 50ms before the "senders'" response to direct stimulation does, suggests one of two possibilities; perhaps the observed effect is not a genuine response to the remote events, as it is not accurately time-locked on these events, but is instead a fluctuation in α-power caused by unknown factors. If however the observed deviation from baseline represents a genuine anomalous response to the remote events, then this would seem to indicate the presence of a temporal as well as a spatial anomaly. As this was not a hypothesis we had considered before the analysis of the results, we can only present this possibility as a question to be explored further in future research. The lack of such a "pre-stimulus" element in responses to direct photic stimulation however, raises the question of why should such anomaly only be present in "remote responses". The temporal asynchrony between direct and "remote responses" (assuming the latter are genuine), would seem to suggest that what we are observing is not an ordinary stimulus-response effect. The physiologically counterintuitive features of the "remote responses" prompt us to suggest that perhaps it is better not to describe these anomalous effects as responses at all, but as "non-local biological interactions" (without implying the involvement of a quantum-mechanical process), or as "remote psychophysiological correlations".

The main limitation of this study was the small number of participants in the "unrelated" and "alone" groups, which made direct statistical comparisons between groups, as well as statistical tests within these two groups impossible to conduct. The related pairs group was the focus of the experiment, and the other two groups were added as exploratory elements within the study. As such they can only be useful for making qualitative comparisons between the groups. We are now planning a larger study with a similar design but with equal numbers of participants in each group, to enable formal statistical comparisons to be made. We are also considering using an "oddball" stimulation paradigm, which could further clarify the physiological characteristics of any anomalous effects found.

ACKNOWLEDGEMENT

This project was funded by the Bial Foundation with grant no. 67/00. The EEG equipment was purchased with the help of a grant from Inova.

REFERENCES

Grinberg-Zylberbaum, J., Delaflor, M., Attie, L., & Goswami, A. (1994). The Einstein-Podolsky-Rosen paradox in the brain: The transferred potential. *Physics Essays, 7*(4), 442-448.

Herz, R. S., Eliassen, J., Beland, S., & Souza, T. (2004). Neuroimaging evidence for the emotional potency of odor-evoked memory. *Neuropsychologia, 42*(3), 371-378.

Jamieson, G. A. (1986). *The structure and meaning of absorption.* Unpublished Masters dissertation, University of Queensland, Australia.

Kawaguchi, T., Jijiwa, H., & Watanabe, S. (1993). The dynamics of phase-relationships of Alpha waves during photic driving. *Electroencephalography and Clinical Neurophysiology, 87*(3), 88-96.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*, 609-621.

Pekala, R. J. (1991). *Quantifying Consciousness: An empirical approach.* New York: Plenum Press.

Radin, D. I. (2003). Thinking outside the box: EEG correlations between isolated human subjects. *Proceedings of the Parapsychological Association, 46*, 184-199.

Sabell, A., Clarke, C., & Fenwick, P. (2001). *Inter-subject EEG correlations at a distance – the transferred potential.* Paper presented at the Proceedings of the 44th Annual Convention of the Parapsychological Association, New York.

Schlitz, M., & Braud, W. (1997). Distant intentionality and Healing: assessing the evidence. *Alternative Therapies, 3*(6), 62-73.

Targ, R., & Puthoff, H. (1974). Information transmission under conditions of sensory shielding. *Nature, 252*, 602-607.

Wackermann, J., Seiter, C., Keibel, H., & Walach, H. (2003). Correlations between brain electrical activities of two spatially separated human subjects. *Neuroscience Letters, 336*, 60-64.

Address for correspondence: Marios Kittenis,
University of Edinburgh, Koestler Parapsychology Unit,
7, George Square, Edinburgh EH8 9JZ, UK.
E-mail: marios@moebius.psy.ed.ac.uk

THE INVISIBLE GAZE: THREE ATTEMPTS TO REPLICATE SHELDRAKE'S STARING EFFECTS

Eva Lobach, & Dick J. Bierman
University of Amsterdam

ABSTRACT

Sheldrake has reported a large number of studies on the feeling of being stared at from behind, mostly conducted in school settings with children aged between 8 and 18. Almost all these studies show that people guess consistently above chance.

We did three studies to attempt to replicate Sheldrake's staring studies under conditions of sensory shielding. Participant input their responses in the computer. In total 188 sessions resulted in 4784 trials. The over-all hitrates were 50.6% (N=53), 52.1% (N=45), and 49.7% (N=37) respectively. Because response bias is believed to confound results for stare and non-stare trials, these are not analyzed separately here, in contrast to Sheldrake's reports.

In study 1 two skeptics and two believers acted as starer. Participants never met the starer who was behind a one-way screen in an adjacent room. Hitrate was marginally significant only for skeptic starers. Extraversion of the participant was irrelevant. It was observed that the most extraverted starer, who was a skeptic, produced the best results.

In study 2 participants were being looked at through a hidden camera. Participants never met the starer. The over-all hitrate of 52.1% was not statistically significant. In this study, skin conductance level of the participant was also used as dependent variable. Initially, when participants were not informed of the staring, skin conductance level showed no difference between stare and non-stare trials. After informing the participants, skin conductance level was shown to be marginally higher during stare trials.

Study 3 compared hitrates for bonded pairs (friends) with non-bonded (strangers) pairs. All participants acted as starer and staree at either side of a one-way screen, once with a friend and once with a stranger. In neither condition, not the slightest sign of a staring effect could be established.

From the results of these three studies we conclude that the staring paradigm is not the easily replicable paradigm that it is claimed to be. Apart from one internal effect all hitrates were very close to chance. Simulations with target sequences identical to those used in earlier studies involving feedback to the starees showed that results of these earlier studies can be explained by a combination of response bias and response strategies rather than psi. Other possible explanations of the failure to replicate concern difference in participants population, type of experimental setting, and rigor of the experimental procedures.

INTRODUCTION

Can we feel when someone is staring at us from behind? Some claim that we can.

One of the most ardent proponents of this claim is Sheldrake, who reports a large number of studies, almost all done with school children (ages 8 through 18) in school settings, showing an apparently repeatable, though as yet unexplained effect.

In a typical study, school children are paired up, one of them looks or does not look at the back of the blindfolded partner in a 20-trial sequence, after which roles of starer and staree are reversed. The starer indicates the start of a trial by a clicker, reponses of the staree (encouraged to be given quickly, often within ten seconds) are noted on a data sheet by the starer (Sheldrake, 2000b).

Responses at stare trials are consistently above chance, while they are around chance at non-stare trials (Sheldrake, 1998, 1999, 2000a, 2001a, 2003a). Sheldrake explains the different results for stare and non-stare trials as a natural phenomenon: we tend to be more aware when some stimulus is present than when it is absent.

A few attempts to replicate Sheldrake's results in more controlled laboratory settings have led to mixed results and to methodological controversies. One of these controversies concerns the possible learning effects of immediate feedback to the staree. According to Colwell, Schröder and Sladen (2000), feedback to starees will not improve their ability to detect an unseen gaze, but will only help to detect patterns in the sequences of stare and non-stare trials. They state that the sequences used by Sheldrake show more alternations (stare/non-stare) than would be expected if the sequences were truly random. When feedback is given, participants will be able to develop an anticipatory strategy that uses a similar alternating pattern as is apparent in the sequences used by Sheldrake. Also, feedback may encourage the staree to try to balance the number of yes and no responses. This strategy will increase the number of hits if the sequence contains an equal number of stare and non-stare trials, as was the case in many of Sheldrake's experiments. Both the alternating and the balancing response strategy may produce results that appear to show a staring effect while in fact no paranormal ability is involved. Sheldrake has since conducted additional studies without feedback, and the reported results are similar (Sheldrake, 2001a). Furthermore, new stare/non-stare sequences are currently in use that try to avoid a balanced number of stare and non-stare trials and the excess of alternating sequences.

One consistent point of critcism is the possibility of sensory 'leakage,' which may allow the staree to use the known senses to detect whether or not he is being stared at. Because the early experiments were done while starer and staree were sitting behind each other in the same room, sensory leakage could not be excluded. In response to this criticism Sheldrake has conducted or supervised a number of studies in which starers and starees are separated by glass windows, and the reported results are still similar (Sheldrake, 2000a).

Response bias of the staree in favor of 'yes' ('I'm being looked at') has also been identified as a potential source of artefactual results. In Sheldrake's experiments participants respond 'yes' in about 55% of the trials. The greater number of hits at stare trials could thus have resulted from this response bias. However, response bias is considered to be a trait of the respondent, independent of condition. This means that hits due to response bias in the stare condition would be balanced out by misses in the non-stare condition. As results in the non-stare condition are at chance level, the over-all results cannot be attributed solely to response bias. Any statement about a *differential* sensitivity at stare *vs.* non-stare trials must be regarded skeptically, however, since these differences are very likely to be obscured by various effects, among them those of response bias and strategic guessing. In a formal description of staring experiments, Van Bolhuis (2004) explored the possibilities of testing the staring effect for stare and non-stare conditions separately under different sets of assumptions. He showed that this could only be done when sensitivity to staring is the sole factor influencing the responses. As soon as the model assumes the presence of any other influences, such as response bias, there is no way to tease apart their respective effects. On the positive side, it can be safely stated that an over-all hitrate is not sensitive to response bias. Since the over-all hitrates reported by Sheldrake are very consistently above chance, they must therefore be caused by 'something else.'

Sheldrake claims that this 'something else' involves a staring effect. The numerous experiments reported by Sheldrake supporting this claim suggest that this staring effect can also be easily replicated. This is rather remarkable, considering that many researchers often find it difficult to replicate psi phenomena, even more so when conscious measures are used as the dependent variable, as is the case in Sheldrake's studies. In the case of the staring effect, laboratory studies using skin conductance variations as the non-conscious dependent variable during stare and non-stare trials have shown less replicable results. Reviewing a number of these studies, the conclusion was that these studies show 'hints of an effect' (Schmidt, Schneider, Utts, & Walach, 2004).

In light of the evidence reported by Sheldrake and his suggestion that these studies make great student projects (Sheldrake, 2003b), we decided to undertake attempts to replicate his findings. Because the three studies reported here were designed and conducted in collaboration with students fulfilling their second-year research project in psychology, they also contain some theoretical extras that reflect the students' interests. In Study 1, we took up the starer's belief in the staring effect and the staree's extraversion as independent variables, in Study 2 we measured skin conductance, and in Study 3 we compared starer-staree pairs of strangers and of friends and fed back the staree's responses to the starer. Common to all studies was the general experimental set up in which starer and staree were placed in separate rooms to avoid sensory leakage as much as we could. Also, no feedback was given to starees, and starees entered their responses directly into the computer.

Although we will report results separately for stare and non-stare trials, we will not conduct separate analyses on these data, because, as we pointed out above, these are heavily influenced by response bias. Instead we report statistical analyses of over-all hitrates only.

To allow for comparison with results reported by Sheldrake, we will also report the percentages of participants scoring more hits than misses, disregarding participants with an equal number of hits and misses. It should be noted, however, that this measure of hits and misses when split for stare and non-stare trials is also strongly affected by response bias, so we will again only test the over-all percentages of participants scoring more hits than misses. Unlike Sheldrake, we will use the binomial p-value, instead of using chi-square. The advantage of the binomial is that it allows for one-sided testing, which is appropriate because the staring effect hypothesis assumes that there will be more participants with hits than misses than vice versa.

If there is a staring effect, obviously there will be more 'yes' responses in the stare condition than in the non-stare condition. Besides reporting over-all response bias, we will also report response bias differences between stare and non-stare trials. We will refer to this difference in response bias as 'possibly psi-mediated response bias.'

STUDY 1

Besides attempting to replicate Sheldrake's results, Study 1 was also designed to explore the effects of the starer's belief in the staring phenomenon and whether the staree's introversion was related to successful staring detection.

Believers are reported to 'produce' stronger staring effects on skin conductance of the staree (e.g. Wiseman & Schlitz, 1997), and are in general found to produce stronger psi effects. The starer's belief was taken up as a within-subjects variable.

Skin conductance of introverts is reported to be more strongly affected by being stared at (Braud, Shafer & Andrews, 1993). We were interested whether introversion was also related to the number of correct responses when explicitly asked to indicate if they felt someone looking at them.

METHOD

Participants

Sixty four first year psychology students participated in exchange for course credit. Students with very high or very low extraversion scores on an earlier administered personality test (5PFT), were especially invited to participate. Others signed up on lists. All participants were informed of the purpose of the study.

Four participants were excluded because they did not cooperate fully, as was apparent from their behaviour, e.g. by reading a magazine during the experiment. Four more were excluded because they answered only 'yes' or only 'no' at all trials in one or both blocks and three more were excluded because they had more than 3 missing values (> 5%). This left 53 participants for data analysis, 39 women and 14 men.

Materials

Rooms. Starer and staree were seated in adjacent rooms, separated by a one-way mirror. In contrast to Sheldrake's studies, participants in this study were seated in such a way that their left side was exposed to the starer's gaze through the one-way mirror. This way, we hoped to make the staring more interesting for the starers.

Starers. Four starers, all women, took turns staring. Two of them believed that people can sense an unseen gaze, and two did not. Each participant was being stared at by a believer (believing starer condition) and a skeptic (skeptic starer condition) consecutively, during one block of 30 trials each, counterbalanced. During a stare trial, the starer looked at the participant without interruption and focused her thoughts on the participant. During a non-stare trial, the starer looked down and tried to think of something else.

Randomization. A visual basic program generated random sequences consisting of an equal number of stare and non-stare trials. Each sequence was printed on a sheet of paper that was consulted by the starer in between trials. A different random sequence was used for each session.

Trial synchronization. To synchronize trials for starer and staree, a taped voice indicated start and number of each trial on both sides of the oneway screen.

Computerprogram. Trials lasted 15 seconds, with a 5 second intertrial interval. At the start of each trial the monitor showed the question "Do you feel that someone is looking at you now?" while a small clock in the corner of the screen counted backwards from 15 to 0 to indicate the time left to enter a response. The response could be given by hitting either one of two keys, meaning 'yes' or 'no'. If no answer was giving during this period, it was registered as missing. During the 5 second interval between trials, the computer presented the question "How certain were you of your answer?" Participants could indicate degree of certainty on a seven point scale ranging from 'completely uncertain' to 'completely certain'.

Extraversion scale. The 14-item extraversion scale of the 5PFT, Cronbach's alpha = 0.85 (Elshout & Akkerman, 1999) measured introversion.

Procedure

One participant was invited at a time. Participants were told that during some of the trials someone would be looking at them from behind the one-way mirror, and that the order of stare or non-stare trials was random. They were instructed to concentrate and to base their answers on intuition and subtle differences in sensations or feeling. Participants did not meet the starer.

After three practice trials, the experimenter left the room and the first block of 30 trials started. After a 5 minute break a second block of 30 trials with another starer was performed. The procedure took between 30 and 40 min.

RESULTS

Of a total of 3180 trials, 36 (1.1 %) were missing because the participant did not respond in time. Hitrate, as reported in Table 1, is corrected for missing trials.

Average hitrates

Average hitrate (Table 1) was significantly different from the expected 50% only in the skeptic starer condition, (52.3%, $t(52)=1.99$, $p=.03$, one-tailed, not Bonferoni corrected), but not in the other condition, nor for the average total hitrates for the two conditions together.

Percentage of participants with more hits than misses

The percentage of participants scoring more hits than misses (Table 1) was never significantly higher than chance level. The skeptic starer condition came closest to statistical significance, exact binomial $p=.15$.[1]

[1] Despite the fact that these percentages are much higher than the overall hitrates, the analysis does not reach significance because much information has been discarded. Instead of an individual hitrate based on 30 trials, a participant contributes only a + or a – to this analysis, which can thus only be tested using Chi-square or Binomial (see also the Introduction).

Table 1: Results of Study 1: Raw hits, hitrates, and percentages of participants with more hits than misses. N=53

CONDITION		# trials	raw hits M (*sd*)	average hitrate	average over-all hitrate	percentage of participants with hits > misses	n[a]	over-all hits>misses	n[a]
Believer	**Stare**	**15**	7.40 (2.39)	49.9%	48.9%	52.9%	51	44.0%	50
	Non-stare	**15**	7.19 (2.24)	48.5%		46.2%	52		
Skeptic	**Stare**	**15**	8.06 (2.14)	54.3%	52.3%	61.5%	52	58.7%	46
	Non-stare	**15**	7.51 (2.27)	50.6%		54.9%	51		
Total	**Stare**	**30**	15.45 (3.53)	52.1%	50.6%	59.6%	52	56.3%	48
	Non-stare	**30**	14.70 (3.82)	49.6%		43.1%	51		

[a] Participants with an equal number of hits and misses are disregarded

Response bias

Over-all response bias in favor of 'yes' ranged from 23% to 86%, average was 52%, not significantly different from the expected 50%, t(52)=1.05, p=.30, two-tailed.

Possible psi meditated response bias was not different in stare and non-stare conditions, except in the skeptic starer condition, M=54% at stare and M=50% at non-stare trials t(52)=1.92, p=.03, one-tailed.

Extraversion, hitrates and response bias

Mean Extraversion score was 62.9 (sd=13.8). We formed two groups of participants around the median (Me=62.5). Over-all staring hitrates were the same for high- and low-extravert participants, as were the percentages scoring more hits than misses.

On average, high-extravert participants responded 'yes' more often. This difference in over-all response bias for high- *vs.* low-extravert participants was statistically significant only in the skeptic starer condition, M=58% *vs.* M=50% respectively. We also observed that the most succesful starer, who generated more hits than the other three in spite of being a skeptic, was the most extravert among the starers.

Discussion Study 1

Contrary to our expectation, the starees who are being looked at by a starer not believing in the effects of staring, appear to detect the staring more accurately than if they are being looked at by a believing starer. This result did not convert our skeptics, though, since it still concerns a very small effect of only 2.3%, not comparable to the larger effects reported by Sheldrake. The effect becomes non-significant (p = .12) after correction for multiple analyses. Note that the non-believing starers involved were no die-hard skeptics and were relatively open-minded about the phenomenon at hand.

Our results show no indication that introverts are better at detecting unseen staring. On the contrary, it seems as if the higher hitrates in the skeptic starer condition are mostly produced by the extraverted starees' greater tendency to say 'yes' when stared at by the skeptics, producing more hits at stare trials. Speculating on the observation that the most successful skeptic starer happened to be the most extraverted of the four starers, it seems possible that the extravered skeptic starer found it easier to bond with extraverted starees (or even vice versa) and that this might have had some effect in the desired direction. Another explanation is that this extraverted starer happened to be very good at staring.

Similarities in personality between starer and staree and skill at staring may be an issue to explore in future studies.

STUDY 2

Considering the over-all rather disappointing results of Study 1, we wondered if our rigourous experimental set-up and the laboratory setting might have interfered with becoming consciously aware of subtle sensations as might be caused by being stared at. Psychophysiological measures, such as electrodermal activity, have on the other hand repeatedly shown to be affected by staring in a laboratory-setting (Braud & Schlitz, 1991; Schmidt et al., 2004). Specifically, Braud and Schlitz (Braud & Schlitz, 1991) showed that skin conductance was lower during stare trials for participants and starers trained in feeling connectedness with each other and the world at large. In contrast, participants not so trained tended to have higher skin conductance when being stared at by a relative stranger.

In Study 2, we attempted to replicate the latter findings using a stranger as the starer, and using a Closed Circuit Television (CCTV) as employed in earlier experiments by Braud & Schlitz (1991). At the same time we wished to explore whether these presumably unconsciously produced physiological effects correlated with conscious responses.

Because informing participants of the purpose of the study might produce unwanted side-effects on skin conductance, the session was split in two blocks. In the first block (naive condition), participants were told that they were to judge the 'tranquility' of musical fragments, while we simultaneously measured their skin conductance level (SCL). During half of the trials, unbeknownst to the participants, they would be stared at through a hidden camera connected to a monitor in an adjacent room. In the second block (informed condition), participants were shown the camera and told that they were being observed during some of the trials. In this second block, instead of judging musical fragments, they had to answer the question whether they felt they were being looked at.

We expected to find higher SCL during stare trials in both the naive and the informed block with an effect size larger than the one obtained by conscious guessing. In addition, we were interested whether variations in SCL correlated with correct guesses. Also, if the participants' answers ('yes' or 'no') correlated with SCL variations, irrespective of stare or non-stare condition, this suggests that their answers might be prompted by these variations.

METHOD

Participants

59 first year psychology students signed up on a list indicating the purpose of the study as 'judging musical fragments', and participated in exchange for course-credit.

Ten participants were excluded because of malfunctioning of GSR-equipment. Four more were excluded because participants answered 'no' at all trials.

This left 45 participants for data analysis, 15 men and 30 women.

Materials

Rooms. Staree and starer were in separate but adjacent rooms. The experimental room contained a relaxing chair in front of a monitor, the UvA equipment to measure SCL, and a response box on which the participants could indicate their answers. A hidden videocamera, behind a binder on a shelf, relayed an image of face and shoulders of the participant in black and white to a monitor in the adjacent observation room. The starer in the observation room sat at the desk that supported the equipment registering the SCL, a monitor showing the screen presented to the participant so the starer would know the number, and start and end of the trial, and a monitor displaying the image of the staree. This monitor was covered by a heavy cloth during the whole procedure, lifted by the starer during stare trials only.

Trials. In both blocks trials lasted 30 s with a 10 s interval in which the participant answered the question presented on the monitor.

Musical fragments. 20 different fragments of soft meditation music were presented through two loudspeakers in trials of 30 seconds each.

Starers. All five experimenters, four women and one man, took turns staring. The starer never left the room and never met the participant.

Randomization. Twenty identical slips of paper, ten with 'stare' and ten with 'non-stare' written on them were shuffled thoroughly and then were picked one by one, noting down their order. This procedure was repeated ten times to create ten different random orders, which were printed out on ten identical sheets of paper. Before each block, the starer picked one of the ten sheets blindly. Before each pick, the sheets were shuffled. If the starer picked the same sequence for the second block, the procedure was repeated.

Computer program. During the first part of the session (naive condition), the computer program indicated the start of the trial by displaying the number of the trial on the staree's monitor and presenting one of the twenty 30 s musical fragments. Each staree heard all twenty musical fragments. During each 10 s inter trial interval, the monitor presented the question: "Did you find this music relaxing?" Participants indicated their response 'yes' or 'no' by pressing a button on the response box with their dominant hand.

The same general procedure was used during the second session (informed condition) but instead the monitor asked: "Did you feel you were being looked at?"

Procedure

The experimenter attached the Ag-AgCl electrodes of the UvA GSR-equipment to the participant's index and middle finger of their non-dominant hand while explaining the procedure. The participant was then left alone for a 10 min acclimatization period after which the experimenter checked if everything was all right and started the experimental program.

After twenty trials the experimenter returned to give instructions for the second block. The whole procedure took about 50 minutes.

RESULTS

Average hitrate

The average hitrate of 52.1 % (Table 2) is higher than chance, but not significantly so, $t(44)=1.24$, $p=.11$, one-tailed. Note that the apparent differential effect between stare and non-stare trials cannot be disentangled from response bias.

Percentage of participants scoring more hits than misses

Testing the percentage of participants scoring more hits than misses (Table 2) against the chance expectation of 50% showed that the binomial probability ($p=.31$) was not statistically significant. As with the average hitrate, analysing results of stare and non-stare condition separately would again have given rise to misleading conclusions.

Table 2: Results of Study 2: Raw hits, hitrates, and percentages of participants with more hits than misses. N=45

CONDITION	# trials	**raw hits** M (*sd*)	**average hitrate**	**average over-all hitrate**	**hits > misses**	**n**	**over-all hits>misses**	**n**
Stare	**10**	4.38 (*2.01*)	43.8%	52.1%	42.9%	35	55.6%	36
Non-stare	**10**	6.04 (*2.28*)	60.4%		73.5%	34		

Response bias

Over-all response bias in favor of yes ranged from 10% to 80%. The average was 42%, significantly different from the expected 50%, $t(44) = -3.07$, $p < .005$ (two-tailed). As in the skeptic condition in Study 1, possibly psi-mediated response bias was higher at stare (44%) than at non-stare trials (40%), but this time the difference was not statistically significant, $t(44)=1.24$, $p=.11$ (one-tailed).

Skin conductance level

In the first, naive, block, mean SCL did not differ in stare (M=18.71, sd=5.79) *vs.* non-stare trials (M=18.73, sd=5.80) as tested in a paired samples t-test, $t(44) = -.61$, *ns*. In the second, informed, block however, participants showed a marginally higher SCL in stare (M=19.26, sd=5.85) *vs.* non-stare trials (M=19.15, sd=5.99), $t(44) = 1.64$, $p = .054$ (one-tailed).

Skin conductance level and conscious verbal response

SCL when the participant said 'yes' as compared to 'no' was on average slightly higher. The average difference was about 1 %, M= .14 (sd=.41), $t(44)=2.15$, $p = .04$ (two-tailed), suggesting that the response to say 'yes' is associated with a higher skin conductance level.

There was no indication that differences in gsr between stare and non-stare trials were related to the number of correct guesses.

Discussion Study 2

When participants were unaware that they were being looked at, their skin conductance level was on average not different between stare or non-stare trials. When they were informed, however, skin conductance was marginally higher at stare trials, as expected. Assuming that no conscious awareness is needed for the skin to respond to an unseen gaze, it is surprising that skin conductance levels only differed between stare and non-stare trials when participants knew about the hidden camera. An explanation might be that knowing about the staring could have made the participant's attention more attuned to subtle variations, resulting in a stronger skin response. Another explanation is that the soothing music in the naive condition, which by necessity always preceded the informed condition, induced a mood that rendered the participant more open to being influenced by the staring. Both ideas could be experimentally tested by varying knowledge about the staring or by letting the actual experiment be preceded by relaxing music or not.

STUDY 3

For Study 3 we communicated elaborately with Sheldrake in order to remove any reasons for our failure to replicate. We were intrigued by Sheldrake's explanation that sensitity to an unseen gaze might have survival value: animals sensing the predator's eyes have more chances to escape and thereby to transmit this sensitivity to their offspring (Sheldrake, 2000c) Although generally people do not relate to each other as predator and prey, we would expect the gaze of a stranger to be more threatening than the gaze of a more familiar person, as was supported by the findings that strangers tend to increase skin conductivity (Braud & Schlitz, 1991), and for which we also found some support in Study 2. In other contexts, Sheldrake suggests that a close relationship or a family bond is particularly advantageous for a staring effect to occur, as was repeatedly shown in experimental findings reported by Sheldrake (e.g. Sheldrake, 2001a). This is also in line with the tentative suggestion from observations made in Study 1, that similarities in personality between starer and staree might create a better bonding and subsequently more hits.

As a first attempt to create more clarity in this matter, we wished to compare success at staring between friends and between strangers. We invited pairs of people with a strong positive bond into our laboratory and asked them to act once as starer and staree together with their partner, and once again as starer and staree with a stranger who was a member of another pair that was invited at the same time. Again, a oneway mirror separated starer and staree, and the starees responded using the keyboard of a computer. During the duration of each trial (10 s) the monitor displayed the image of a flower.

The procedure is similar to the procedure used by Sheldrake in the sense that normal participants acted both as starer and staree. To make the staring role more interesting, we decided to provide feedback on the responses and on the over-all hitrate of the staree. This allows the starer to check the effectiveness of various staring techniques (*cf*. Braud & Schlitz, 1991).

METHOD

Participants

Forty-four participants were recruited in pairs of good friends or relatives among family and friends of the experimenters in exchange for 7 Euros. Two pairs participated at the same time in a four-part session. All participants took turns as starer and staree, once with the partner they signed up with (*friends* condition), and once with a member of the other pair (*stranger* condition), counterbalanced.

Because one of the participants became unwell during a session, the data of the four participants in that session were lost. Three more were excluded because they only answered ‘no’ in one or both blocks. This left data of 37 participants for analysis, 10 men and 27 women.

Method

Rooms. Two adjacent oneway-mirror units were used simultaneously. Starer and staree were both seated in front of a monitor. Both monitors displayed the same image at all times, but the bottom part of the screen, showing instructions to the starer, was hidden from the staree behind a strip of paper taped to the monitor. The starer’s monitor was facing perpendicular to the one-way screen in such a way that the starer would only have to glance sideways to check whether it was a stare or non-stare trial, and no movement of the chair was necessary to adjust posture from a stare to a non-stare trial or vice versa.

Computerprogram. To synchronize the procedure at both sides of the screen, the start of each 10s trial was indicated by a loud beep that was also audible on the other side of the one-way mirror. At the beep a picture of a flower was presented and stayed on screen for the duration of the trial, followed by the request: “Give your impression if you are being looked at now,” together with a representation of the four response options “Yes I feel I am being looked at and I am certain,” “Yes I feel I am being looked at and I am uncertain,” “No, I don’t feel I am being looked at and I am uncertain,” and “No, I don’t feel I am being looked at and I am certain.” The staree indicated the appropriate answer by pressing any of four keys representing the four options during the 10 s intertrial interval.

At the bottom of the starer’s monitor the type of trial (stare or non-stare) was indicated directly at the beep indicating the start of the trial. It was replaced by feedback about the response of the staree and cumulative number of hits during the intertrial interval.

Randomization. The computer presented twenty[2] trials (ten stare and ten non-stare) in a random order, different for each participant.

Procedure

Two pairs of participants were invited at a time. Each pair was assigned to one of the two experimenters to guide them to their rooms and give instructions throughout the session. A session consisted of four parts.

[2] Ten more trials were presented in a fixed order as the middle ten of a total of 30 trials to test another hypothesis that we will not discuss here. They are left out of the analyses presented here

Each participant acted as starer and staree, once together with a friend (part one and two) and once together with a stranger (part three and four). In half of the sessions pairs started with the stranger condition; pairs were split up and coupled to a member of the other pair. In the other half of the sessions, pairs first worked together as starer and staree. A toss of a coin determined who would be starer first.

The staree was seated in front of a monitor, his back to the one-way mirror. The starer was seated behind the mirror. The participants were given print-outs of the instructions and allowed some time to read. Besides explaining the procedure, the instructions advised both starer and staree to focus on the image on the screen and to relax. After three practice trials the experimenters left the participants. During the second part of the session, the experimenters let starer and staree switch places and handed them instructions on their new roles. During the third and fourth part of the session, new pairs were formed. At the end of the session, participants filled out a short questionnaire asking them about their belief in the phenomenon, about their current feelings towards the friend they had signed up with, and whether they knew the member of the other pair they were paired up with for the stranger condition.

RESULTS

Hitrate and percentage of participants scoring more hits than misses

In none of the conditions was the hitrate higher than chance level, and neither was the percentage of participants scoring more hits than misses. (Table 3)

Table 3: Results of Study 3. Raw hits, hitrates, and percentages of participants with more hits than misses. N=37

CONDITION		# trials	raw hits M (*sd*)	average hitrate	average over-all hitrate	hits > misses	n	over-all hits>misses	n
Friends	**Stare**	**10**	5.00 (*1.58*)	50.0%	48.9%	42.4%	33	47.1%	30
	Non-stare	**10**	4.78 (*1.65*)	47.8%		50.0%	28		
Strangers	**Stare**	**10**	4.73 (*1.54*)	47.3%	50.4%	32.1%	28	53.3%	33
	Non-stare	**10**	5.35 (*1.70*)	53.5%		55.6%	27		
Total	**Stare**	**20**	9.73 (*2.64*)	48.7%	49.7%	40.0%	35	47.1%	34
	Non-stare	**20**	10.14 (*2.64*)	50.7%		40.0%	30		

Response bias

Response bias in favor of yes ranged from 25% to 70%, average was 49%, not significantly different from the expected 50%, t(36)= -0.36, *ns*. There was no indication of a possibly psi-mediated response bias; response bias differed for stare and non-stare trials in none of the conditions. However, participants gave on average more 'yes' reponses in the friends condition (M=51%) than in the stranger condition (M=47%), a statistically significant difference, t(36)= 2.19, p=.04 (two-tailed).

Discussion Study 3

Results in Study 3 not only failed to establish whether the gaze of a friend can be detected more easily than the gaze of a stranger, it also failed to produce any indication that people are sensitive to being stared at from behind. One possible explanation for this negative result is a somewhat unfortunate aspect in the design that we only realised while past midway conducting the experiment: all participants saw all images of flowers in exactly the same order for all four parts of a session. When the participant acted as starer, in either the first or the second part of a session, this could easily create an association between a particular image and the type of trial, stare or non-stare, which might prompt an answer during the next part when the participant acted as staree. Assuming this is the case, only the first part of each session would be free of this kind of association. Testing hitrates from these first parts of the sessions, however, did still not show any indication of a staring effect.

Another possible artifact was the position of the starer behind the one-way screen. Because the experimenters were not present to check if starers followed strictly protocol, it is possible that their position allowed sideways glances through the screen. Although this could have happened, it seems rather unlikely that this would have happened very often or out of carelessness or lack of commitment, since almost all participants were good friends or close relatives of the experimenters and were in general highly motivated to collaborate and follow instructions carefully. Nevertheless, accidental sideways glances could have 'contaminated' the non-stare trials and thereby diluted a potential staring effect.

GENERAL DISCUSSION

The three studies presented above show a rather disappointing picture as far as the replicability of the staring effect is concerned. None of the studies found the large staring effects reported by Sheldrake. The straightforward conclusion must be that staring effects are not easily replicable, although they do indeed serve as great student projects.

We are left with the question of how to explain the discrepancy between Sheldrake's results and ours. Several answers spring to mind.

An important difference was the use of feedback in many of Sheldrake's earlier staring studies. It seems possible that a combination of response bias in favor of yes and guessing strategies in response to feedback could explain his results. In order to test this possibility, we ran a simulation using the stare/non-stare sequences as published at Sheldrake's website, *http://www.sheldrake.org*. There are two kinds of target sequences. The old 24 target sequences were created in response to recommendations by Wiseman and Smith (1994). These old sequences contained an equal number of stare and non-stare trials, and 12 sequences were the mirror image of the other 12. Criticisms by Colwell et al. (2000) about the large number of alternations prompted the creation of 20 new sequences that are all different, do not have a balanced number of stare and non-stare trials and do not show as many alternations as the old sequences.

We ran three different simulations on both the old and the new series of sequences, using 100 artificial subjects for each target sequence. The first was a mere chance simulation, without response bias and without guessing strategy, the second used a response bias of 55% in favor of 'yes', and the third used a response bias of 55% together with a guessing strategy in response to feedback.

The guessing strategy assumed that the participant adjusts his guessing probabilities ('yes' and 'no') to the feedback on the previous trials. This guessing strategy, which is akin to a gambler's fallacy, can be expressed (without response bias) as the probability to say 'yes' after feedback of n_stare + n_nostare trials:

p_say_yes_WB = (10 - n_stare) / (20 - n_stare - n_nostare)

where p_say_yes_WB is the probability the participant says 'yes' (without response bias), n_stare is the number of times that according to the feedback the subject was stared at and n_nostare the number of times the subject was not stared at. At the beginning of the session no information is available, n_stare and n_nostare are of course zero, and the probability to say 'yes I am being looked at' is 0.5.

For the third simulation this guessing strategy was combined with a response bias of 55% in favor of 'yes'. The strategy used in this simulation can be expressed as:

p_say_yes_B = p_call_looking_WB + bias

Table 4: Hitrates in a simulation with Sheldrake's 24 old and 20 new target sequences with No Feedback & No Response Bias, with No Feedback & Response Bias 55% and with Feedback & Response Bias 55%.

Simulation condition		Old target sequences (n=2400)	New target sequences (n=2000)
		hitrate	**hitrate**
No feedback – No bias	**Stare**	50.7%	49.9%
	Non-stare	49.6%	48.6%
No feedback – Bias 55%	**Stare**	54.9%	55.1%
	Non-stare	45.1%	43.8%
Feedback[a] – Bias 55%	**Stare**	59.1%	59.0%
	Non-stare	49.5%	44.7%

[a] Simulation with feedback assumes that participants use a gambler-fallacy type of strategy. See text.

The results of these simple simulations are given in Table 4. We can see that the results of the mere chance simulation, without strategy and without response bias, look quite normal. This is true for the hitrates in both stare and non-stare condition. The simulation with a response bias of 55% shows the expected above chance hitrate in the stare condition, which is compensated by about the same under chance scoring so that the over-all results are at chance.

The most interesting results are those of the third simulation, using the above guessing strategy in combination with a response bias. The results thus obtained for the old (balanced) sequences mimic quite closely the results reported in the literature: they are at chance in the non-stare condition while those in the stare condition are around 60%. This of course can be expected given the balanced nature of the old target sequences. To our surprise, however, the very same strategy also yielded artificial psi scoring for the new sequences. That is, if one pools both conditions there is still a clear above chance scoring percentage of about 51.8%, statistically extremely significant. This unexpected result suggests that even among the new target sequences there are still too many that are close to a balanced sequence. On inspection, 16 of the 20 new target sequences have a 10-10 or a 9-11 distribution, and only 4 have a 8-12 distribution.

These simulations show that the effects found in staring studies involving feedback to the staree might explain the difference between the effects reported by Sheldrake and the absence of those effects in ours.

However, as we noted above, Sheldrake has since reported quite a number of studies that do not provide feedback to the staree, and these show somewhat smaller, but still large effects (Sheldrake, 2001b), unlike our three studies, so it leaves our question in part still unanswered.

Another important difference is that Sheldrake's experiments are almost exclusively conducted with school children from age 7 or 8 and up, while our experiments were mostly conducted with young adults, or older. However, Sheldrake has reported a number of studies with adults as well, with similar results.

Another difference is the use of blindfolds. We never used blindfolds in our studies, while Sheldrake has adopted the use of these in many of his staring experiments. It seems possible that blindfolds help participants to focus on the presumably subtle sensations involved in staring.

Still another difference is the experimental setting: the familiar environment of the school *vs.* a relatively sterile laboratory setting. Although the familiar school environment might provide the relaxed atmosphere that promotes psi phenomena, it also makes it much harder to adhere to controlled circumstances, which points to still another difference between our experiments and those of Sheldrake. In the latter the scoring is done by hand, and there are no double scoring sheets which would allow for double checking. One scoring error per sheet in favor of the staring hypothesis would suffice to get the results that Sheldrake reports. Although it is extremely unlikely that so many scoring errors have been made, adopting either independent hand scoring or automatic registration with full date and time information would prevent this concern. This way, the best of both approaches could be combined. The results of the laboratory studies that we conducted so far, leave no other conclusion than that the conscious effects of an unseen gaze remain invisible.

References

Braud, W. & Schlitz, M. (1991). Consciousness interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies, 2-1,* 1-46

Braud W., Shafer D., & Andrews S. (1993). Further studies of autonomic detection of remote staring: replication, new control procedures, and personality correlates. *The Journal of Parapsychology, 57,* 391-409.

Colwell, J., Schröder, S., & Sladen, D. (2000). The ability to detect unseen staring: A literature review and empirical tests. *British Journal of Psychology*, 91, 71-85.

Elshout, J.J., & Akkerman, A.E. (1975). *Vijf persoonlijkheidsfaktoren test, 5PFT [Five personality factors test, 5PFT],* Berkhout, B.V. – Nijmegen, The Netherlands.

Schmidt, S., Schneider, R., Utts, J. M. & Walach, H. (2004). Distant Intentionality and the Feeling of Being Stared At - Two Meta-Analyses. *British Journal of Psychology, 95, 235-247.*

Sheldrake R. (1998). The Sense of Being Stared At: Experiments in Schools. *Journal of the Society for Psychical Research, 62,* 311-323.

Sheldrake, R. (1999). The sense of being stared at confirmed by simple experiments. *Biology Forum, 92,* 53-76

Sheldrake, R. (2000a). The „sense of being stared at“ does not depend on known sensory clues. *Biology Forum 93* 209-224.

Sheldrake, R. (2000b) Research on the sense of being stared at. Available: http://www.sheldrake.org/experiments/staring/staring_experiment.html

Sheldrake, R. (2000c). *Dogs that know when their owners are coming home and other unexplained powers of animals.* Arrow Books – London, United Kingdom.

Sheldrake, R. (2001a). Experiments on the sense of being stared at: The elimination of possible artefacts. *Journal of the Society for Psychical Research, Vol. 65,* 122-137

Sheldrake, R. (2001b). Follow-up. Research on the feeling of being stared at. *Skeptical Inquirer magazine, march/april* Available: *http://www.scicop.org/si/2001-03/stare.html*

Sheldrake, R. (2003a). *The sense of being stared at and other aspects of the extended mind*, Hutchinson – London, United Kingdom.

Sheldrake, R. (2003b). Lecture given as part of the Sentient Creatures Series at the Waag, Amsterdam, The Netherlands, 8 October 2003. Available: http://connected.waag.org/archive.html

Van Bolhuis, J. (2004). An analysis of response bias in some of Rupert Sheldrake's experiments. *Unpublished document,* Free University, Amsterdam, The Netherlands.

Wiseman, R. & Schlitz, M. (1997). Experimenter effects and remote detection of staring. *Journal of Parapsychology, 61*, 197-208.

Wiseman, R. & Smith, M. (1994). A further look at the detection of unseen gaze. *Proceedings of the Parapsychological Association 37th Annual Convention.* Parapsychological Association. 465-478.

Address for correspondence: Dick Bierman,
University of Amsterdam, Roetersstraat 15,
1018 WB Amsterdam, The Netherlands.
E-mail: d.j.bierman@uva.nl

WHO'S CALLING AT THIS HOUR? LOCAL SIDEREAL TIME AND TELEPHONE TELEPATHY

Eva Lobach, & Dick J. Bierman
University of Amsterdam

ABSTRACT

Can we guess who is calling us on the phone before picking up, and does local sidereal time (LST) affect how often we guess right? Reviews of anomalous cognition studies have shown that effect sizes are highest around 13.30 LST (Spottiswoode, 1997). A post-hoc analysis of telephone telepathy data of Sheldrake (2003) also showed a peak at that time. LST (peak or non-peak) was an independent variable in our prospective telephone telepathy study. Six women who indicated they often experienced telephone telepathy were selected to participate. Each participant chose four close friends or relatives to act as callers. All completed a total of 36 trials; six sessions of six trials each, three sessions at peak time (between 8.00 and 9.00 local time) and three at non-peak time (between 17.30 and 18.30 local time). One of the experimenters was at the participant's home during the sessions. The experimenter made sure no irregular communication was going on and logged times of the calls and responses of the participant. At a different location another experimenter used a dice to select a caller about five minutes before the scheduled trial. Then he or she contacted the caller who was asked to call the participant in five minutes and to concentrate his or her thoughts on the participant for the last two minutes before the call was made. When the phone rang at the participant's home, the participant guessed who she thought was calling before picking up. Analyses show a significant over-all scoring rate of 29.4% (p = .05). Almost all of this effect originates from the sessions at peak time with a scoring rate of 34.6%. Exploratory analyses show that a stronger emotional bond between particpant and caller is associated with a higher hitrate. It is concluded that results provide tentative support for the hypothesis that Local Sidereal Time is related to a phenomenon like telephone telepathy. In addition, the results are in support of the existence of telephone telepathy. Other explanations of the anomalous effect cannot be ruled out, such as precognition, retro psychokinesis by the experimenter or the participant so the dice throw would coincide with the particular caller the participant would guess, or clairvoyance of the dice throws. Future studies should aim at teasing apart the supposed effects of LST and local time on 'telephone telepathy.'

INTRODUCTION

You are at home. The phone rings. And you know who is calling before answering. This kind of 'telephone telepathy' appears to be an ordinary phenomenon, although some of us appear to be better at it than others. A number of studies with pre-selected individuals who performed above chance in a pilot study showed evidence that some people are quite good at guessing who is calling them, especially when the callers are good friends or relatives (Sheldrake, 2003).

Interestingly, there is evidence that telephone telepathy, which is considered to be a form of anomalous cognition (AC), not only varies among individuals, but also varies along the hours of the day, or more accurately, by local sidereal time (LST). In a large sample of anomalous cognition studies, Spottiswoode (1997) found that effect sizes were highest around 13.30 h LST. Although there is no obvious explanation for such a finding, the same distribution of effect sizes was confirmed in a separate sample of anomolous cognition studies. Submitting Sheldrake's telephone telepathy data to the same procedure, these again showed the largest effect sizes around 13.30 h LST (Spottiswoode, 2003).

These findings are very suggestive, but they are post-hoc analyses, and therefore a prospective study is much needed. Our goal was to replicate the telephone telepathy study of Sheldrake, and use LST as a within-subjects independent variable.

Although we implemented a few changes in order to obviate criticisms aimed at Sheldrake's (2003) studies, the general lay-out of our study was very similar. Participants were at home, since a familiar environment is assumed to foster telepathic sensitivity, and would receive several phone calls on their landline phone (without number display), in a previously arranged period of time. The participant knew that the dice thrown by the experimenter at some different location would determine which one of four different callers would be calling. Before picking up, the participant would state clearly who it was.

Instead of audio- or videotaping the sessions as Sheldrake did, we chose to have one of the other experimenters visit the participant, in order to check if there were no other types of communication going on before the phone calls were being made so as to preclude any suspicions of hidden accomplices.

All participants took part in six sessions, three at peak LST (peak condition), around 13.30 h LST, and three at non-peak LST (non-peak condition) around 00.00 h LST. Participants were not informed of the LST hypothesis.

Although Spottiswoode's analyses showed that the lowest AC effect sizes occurred around 18.00 h LST, we chose 00.00 h LST as our non-peak time for a pragmatic reason. When this experiment was conducted (December 2003 / January 2004, Amsterdam, The Netherlands), peak LST was between 8.00 and 9.00 AM, and Spottiswoode's low at 18.00 LST would be around lunch time. We reasoned that it would be difficult to organize sessions at the participants' homes around that time, and instead chose to conduct the non-peak sessions around 00.00 LST, which was between 6.30 and 7.30 PM local time. According to Spottiswoode's analyses, this would be a time of about average effect sizes.

METHOD

Participants

Participants were recruited through e-mails sent to friends and acquaintances of the experimenters. They had to meet four requirements: they had to be available at the scheduled times, live close enough to one of the experimenters, find four good friends or relatives to cooperate, and have experienced that they often guessed correctly who was calling them on the phone.

Six women, aged 16, 19, 20, 21, 21, and 54 were selected. They received €40 for taking part in all six sessions. The callers, four for each participant, 24 in total, each received €15. Participants were at home during all sessions. Callers could be anywhere, as long as they could be reached by phone.

Materials

Local Sidereal Time. Local sidereal times were calculated by means of on online scheduler, Local Sidereal Time – Monthly experiment scheduler for PSI research (Melssen, 2003).

Number of sessions and trials. Each participant took part in six sessions, three at peak and three at non-peak LST. Each session consisted of six trials, so there were about ten minutes in between the calls. Participants were each called 36 times, the total number of trials was 216.

Emotional bond. Before the start of the first session, the participants indicated their emotional bond with each of the four selected callers on a five point scale from 1, 'good emotional bond', to 5, 'very strong emotional bond'.

Logs. There were two logs, one was kept by the experimenter visiting the participant ('visiting experimenter'), and the other one was kept by the experimenter selecting the callers ('control experimenter'). Both logs bore the name of the participant and names of the callers, date, LST condition and session number.

The control experimenter wrote down the time at which each call was made to the selected caller, and who was selected to make the call. Any irregularities, such as when a caller couldn't be reached, were also logged.

The visiting experimenter wrote down the time the participant received the call, the response of the participant and whether it was a hit or a miss, according to the participant.

Randomization. A number between one and four was pre-assigned to each caller. A dice thrown by the control experimenter determined which caller had to make a particular call. Five and six were disregarded.

Procedure

Just before the scheduled time, the visiting experimenter arrived at the participant's home and checked if everything was in order. He or she then retreated to a different room, within hearing distance of the participant, but out of sight, so as to disturb the home environment of the participant as little as possible.

The callers weren't at any specific location, but were asked to be available by phone at the scheduled times.

At the start of each session, the control experimenter contacted the four callers and the participant to tell them that the session had started and to check if everything was ready at the participant's home. If it turned out that not all callers were available at the start of the session, the procedure was still followed through.

Before each trial, the control experimenter threw a dice to select the caller for that trial. The caller was told to call the participant in five minutes, and to concentrate his or her thoughts on the participant for the last two minutes before making the call. If the caller's line was busy or if the caller didn't answer, the control experimenter contacted the caller again in three minutes and asked the caller to call the participant in two minutes. If the line was still busy or the caller was still unavailable, the experimenter would throw the dice again to select another caller.

When the phone rang at the participant's home, the participant said who she thought was calling, and the response was logged by the visiting experimenter. The participant then answered the phone and told the visiting experimenter whether she was right or wrong.

After six trials the control experimenter contacted the callers to tell them that the session was over, to thank them for their cooperation and to remind them of the next session.

RESULTS

Session integrity

Logs of the two experimenters were put together in one datafile. In two cases we found a discrepancy between the two logs. According to the logs, the participant's guess and the caller were the same, so it should have been a hit, but the visiting experimenter had logged it as a miss. These two trials were further considered as missing values.

The logs showed that in ten of the 36 sessions, one or more of the callers were partly or totally unavailable for various reasons. They could have missed the appointment entirely, or could have forgotten to turn there mobile on in time. In a number of cases, a caller was not available during part of a session, e.g., because she was taking a shower. It seemed unlikely, though not impossible, that the participant would know about the absence of one of the callers. The chances of guessing correctly if one is aware of the absence of a caller is 1/3 rather than the a priori hit probability of 1/4. Therefore we decided to analyse the data twice: once with all sessions included (Table 1), and once again with only the regular sessions (Table 2).

Distribution and randomness

To acquire an indication of the 'randomness' of the dice throws and the participants' guesses, we checked the frequency distributions for all trials and all consecutive pairs of trials.

The frequency distribution of dice throws and of the participants' guesses did not differ from the expected 25% for each of the four options, χ^2 (3) = 3.53, p > .3 and χ^2 (3) = 4.51, p > .2 respectively.

As an indication of the randomness of the sequences, we looked at all pairs of consecutive dice throws and guesses. If the sequences are random, consecutive pairs of similar numbers (in this case 11, 22, 33, and

44) are expected to make up 25% of the total of all consecutive pairs, while 75% of the pairs would be dissimilar.

For the dice, 35 (20%) of the 179 consecutive pairs contained similar numbers, less than expected according to chance. The difference with chance expectation was not statistically significant, however, χ^2 (1) = 2.83, p =.09.

For the guesses, 52 (29%) of the 179 consecutive pairs contained similar numbers, more than expected according to chance, but again the difference was not statistically significant, χ^2 (1) = 1.47, p = .23.

Although these tests cannot firmly establish whether the sequences are truly random, it should be noted that the deviations from chance expectation for dice throws and guesses are in opposite directions. This means that it is highly unlikely that peculiarities of the sequences per se could account for an above chance hitrate.

Over-all hitrates

Table 1: Number and expected number of hits at peak LST sessions, non-peak LST sessions and total

	PEAK			**NON-PEAK**			**TOTAL**		
PARTICIPANT	trials	hits	H_{exp}	trials	hits	H_{exp}	trials	hits	H_{exp}
1	18	8	4.50	18	4	4.50	36	12	9.00
2	18	5	4.50	18	4	4.50	36	9	9.00
3	18	6	4.50	18	2	4.50	36	8	9.00
4	17	4	4.25	18	6	4.50	35	10	8.75
5	18	9	4.50	17	4	4.25	35	13	8.75
6	18	5	4.50	18	6	4.50	36	11	9.00
total	107	37	26.75	107	27	26.75	214	63	53.50

Notes: Total number of sessions was 36, 18 at peak and 18 at non-peak LST. Each participant took part in 6 sessions (3 at peak and 3 at non-peak LST), and received 6 calls per session. For two trials, one peak-trial for participant 4 and one non-peak trial for participant 5, the logs of the experimenters showed discrepancies. These trials are therefore registered as missing values. H_{exp} indicates the expected number of hits.

Pooling peak and non-peak conditions together, the over-all scoring rate is 29.4% with all sessions included (Table 1) and 32.7% for the subset of sessions that confirmed strictly to the protocol (Table 2). Testing individual scoring rates of the participants against expected scoring rates (25%), a paired-samples t-test shows that scoring rates are significantly above chance, t(5)=2.01, p=0.05 (one-tailed) when all sessions are included (Table 1), and also for the subset of the 26 regular sessions (Table 2), t(5)=5.48, p<.005 (one-tailed).

Table 2: Number and expected number of hits at peak LST sessions, non-peak LST sessions and total for the 26 sessions where callers were available throughout

	PEAK			**NON-PEAK**			**TOTAL**		
PARTICIPANT	trials	hits	H_{exp}	trials	hits	H_{exp}	trials	hits	H_{exp}
1	12	6	3.00	18	4	4.50	30	10	7.50
2	18	5	4.50	12	4	3.00	30	9	7.50
3	12	4	3.00	~	~	~	12	4	3.00
4	12	3	3.00	18	6	4.50	30	9	7.50
5	18	9	4.50	12	2	3.00	30	11	7.50
6	18	5	4.50	6	3	1.50	24	8	6.00
total	90	32	22.50	66	19	16.50	156	51	39.00

Note: Results are based on 15 peak sessions and 11 non-peak sessions, a total of 26 sessions. See also the notes at Table 1.

Telephone telepathy and LST

As Table 1 shows, the scoring rate was higher than expected in the peak condition (34.6%), while it was around chance (25.2%) in the non-peak condition. To test the statistical significance of this 9.4% difference, we compared scoring rates in peak and non-peak condition for each participant. A paired-samples t-test showed that scoring rates were marginally significantly different between peak and non-peak condition, $t(5)=1.60$, $p=.09$ (one-tailed).

For the sessions in which all callers were available throughout (Table 2), the average scoring rate was 35.6% for the peak and 28.8% for the non-peak condition. This time, we could only compare five participants, because all non-peak sessions of one participant had to be dropped from analysis. The difference of 6.8% between peak and non-peak condition was not statistically significant.

Emotional Bond and Hitrates

The average emotional bond with the callers was $M=3.71$ ($sd=1.12$), indicating that the participants had indeed a strong emotional bond with the callers. To explore the extent of response bias due to emotional bond, we calculated the correlation between emotional bond and the number of times a participant guessed a particular caller. This correlation was performed using all 214 trials, after normalizing the variables, and turned out to be not significantly different from zero, $r=.20$, *ns*.

To explore if emotional bond between participant and caller was related to success at guessing, we again first normalized the emotional bond ratings and scoring rates for each participant separately, including all 214 trials. The correlations between emotional bond and scoring rate ranged from –0.73 to 1.00. Five of the six participants showed a positive correlation. Overall correlation between emotional bond and scoring rate was $r=.41$, $p<.05$. Controlling for response bias (the number of times the participant guessed a particular caller), the correlation between emotional bond and hitrate still reached about the same value, $r=.39$, $p=.07$.

DISCUSSION

Results show weak support for the hypothesis that Local Sidereal Time affects an anomalous cognition phenomenon like telephone telepathy. Our sample was rather small, and the large number of sessions that showed irregularities further weakened the power of our statistical analyses. However, the results are still promising and replications are warranted.

Future studies should take into account that this single study could not separate the effects of local time and local *sidereal* time. It is possible that the tentative support for the LST hypothesis must be attributed to the fact that the LST peak sessions were all conducted early in the morning. Although we are not aware of any evidence that anomalous cognition effects are related to the time of day, it cannot be excluded. When the LST peak times are during the evening, however, it will always be summer at this location of the globe. Type of season is another factor that could potentially affect telephone telepathy. Replications in other continents are therefore especially helpful to gather more evidence concerning the effect of LST on anomalous cognition.

Results are more supportive of the existence of telephone telepathy. Overall, our participants guessed correctly more often than would have been expected according to chance. The evidence is not as impressive as the 45% results reported by Sheldrake (2003), however, which may have been due to his preselection of participants through a pilot study, while we employed a rather loose selection criterion by including only those participants who reported to have had experiences of telephone telepathy in the past.

The correlation between emotional bond and hitrate in our study showed that even in a relatively homogeneous group of friends and relatives, people still appear to be better at guessing those callers with whom they have the strongest emotional bond. This result is in line with Sheldrake's findings.

The effect size reported here is very close to the typical effect size found in Ganzfeld experiments. Thus the required number of trials from a power perspective (power=0.5) is between 200 and 250 to obtain a result significant at the 5% level. However, there is a huge difference in time investment to obtain this result. We ran 12 trials per day while in Ganzfeld research it is typically 1.

Could the telephone telepathy in this experiment be explained by more 'ordinary' circumstances during the experiment?

There are some concerns.

Logs were kept by hand and experimenters could have made errors. In fact, we identified two errors (in a total of 214 trials), where the response of the participant and the name of the caller matched, but the log of the visiting experimenter nevertheless registered a miss. We treated these two instances as missing trials. Although there may have been one or two additional but undiscovered mistakes in registering the name of the caller or the actual response of the participant, it seems highly unlikely that a hit or a miss was logged inaccurately. There were only six trials in an hour, and both the participant and the experimenter were very keen on the outcome of every trial.

Could the callers have notified the participant before the call was made? This seems possible, using mobile phones in the buzz mode. A potential caller could have a code of 1, 2 or 3 buzzes before hanging up. The visiting experimenter was aware of this option, however. Furthermore, there was not much time left for the caller between being contacted by the control experimenter and making the actual call to the participant.

The participant could know some callers to be late, and this might produce subtle timing differences between callers. We had discussed the adoption of a random timing procedure, so the control experimenter would contact the callers at varying time intervals, instead of keeping to a ten-minute interval, as was the case now. However, we suspected that the timing of trials would be subject to some fluctuations anyway, e.g., when it would take the experimenter longer to throw a number under five, or when the caller was not directly available. Nevertheless, timing differences between callers might still provide an explanation of correct guesses at some trials, and future studies should consider a random timing procedure, in combination with a less cumbersome randomization procedure.

Lastly, there might have been a collusion between student experimenters and student subjects. Only replications with extra technical security measures might be able to make this type of explanations less acceptable.

It seems, nonetheless, that the results of this study provide further evidence for the existence of some anomalous effect, but should we call it 'telepathy'? In spite of using instructions for the caller to concentrate and think about the person to be called, instructions consistent with the telepathy model, some other options are possible.

First, there could be precognition at work. Precognition appears to be dependent on the time between the guess and the feedback and in this study, this time was short and hence could have produced good precognition results. Second, psi could have entered at the experimenter randomization level. That is, the dice throwing was such that it correlated with what the participant would say anyway. This could be seen as retro psychokinesis by the experimenter or the participant. And finally, it could be clairvoyance: The participant could be 'tuned' to the dice being thrown by the experimenter and know who was selected for the call as soon as 'the die was cast.' Considering this option, we inspected the trials in which the dice indicated a particular caller who later turned out to be unavailable. This happened at eight trials. In five of these eight trials, the participant guessed the name of the caller who should have called if he had been available (exact binomial $p=.03$). Of course, instead of clairvoyance, the 'missed hits' could also indicate telepathy with the control experimenter. Note that the finding that participants still mentioned the names of callers who turned out to be unavailable, counters the suggestion that 'psi' effects could have resulted because the participant was somehow informed of the absence of one of the callers.

In light of the generally unsuccessful replications of psi research at the university of Amsterdam the current results should be considered as a recommendation for further exploration of the telephone telepathy paradigm.

References

Melssen, F. (2003). Local Sidereal Time – Monthly experiment scheduler for PSI research. *http://oase.uci.kun.nl/~owc/cgi-bin/lst/2/lstcal3.pl*.

Sheldrake, R., & Smart, P. (2003). Videotaped experiments on telephone telepathy. *Journal of Parapsychology, 67,* 187-206.

Spottiswoode, S.J.P. (1997). Apparent association between effect size in free response anomalous cognition experiments and local sidereal time. *Journal of Scientific Exploration, 11,* 1-17.

Spottiswoode, S.J.P. (2003) "Telephone telepathy." Analysis of data for LST and GMF effects. Private communication.

Address for correspondence: Dick Bierman,
University of Amsterdam, Roetersstraat 15,
1018 WB Amsterdam, The Netherlands.
E-mail: d.j.bierman@uva.nl

PREDICTIONS OF THE MODEL OF PRAGMATIC INFORMATION ABOUT RSPK

Walter von Lucadou & Frauke Zahradnik
Parapsychological Counseling Office, WGFP, Freiburg

ABSTRACT

The model of pragmatic information (MPI) is applied to RSPK phenomena and leads to several predictions. The first prediction is that RSPK phenomena show two clusters of phenomena, which can be considered as structural and functional anomalous RSPK- phenomena. RSPK-phenomena are considered as a kind of "external psycho-somatic" reaction, expressing a hidden problem, which cannot be recognized by the persons concerned. The second prediction is that the development of RSPK cases contains four phases, which are called "surprise phase", "displacement phase", "decline phase", and "suppression phase". In the surprise phase the RSPK-activity starts rapidly with strong effects, but they are not attributed to the focus person. This happens in the displacement-phase where the phenomena usually change in an unpredictable way. In the decline-phase the "message of the poltergeist" is understood and the phenomena are expected, therefore the phenomena disappear. The final suppression phase can be understood as a kind of reaction of the society. These phases can be derived from a fundamental equation of the MPI, which describes the RSPK phenomena in complementary terms of "autonomy" and "reliability" (from the point of view of the RSPK-system) and of "novelty" and "confirmation" (form the point of view of the external observer). The dynamics of RSPK is described as the dynamics of pragmatic information within a hierarchically nested system, which is created by the persons involved (focus person, naïve and critical observers) and the reaction of the society. The third prediction is that observers can control the RSPK activity by their observation or documentation. This is the result of a kind of "uncertainty relation" of the MPI, which says that the effect-size of the phenomena is limited by the quality of their documentation. This also holds for so-called sitter-group experiments. In a single case it was possible to demonstrate that the system theoretical approach of the MPI leads to different predictions than the usual psychological interpretation of the elusiveness of PK-phenomena. The fourth prediction is that we have to expect two types of RSPK cases, which we called the active RSPK case and the passive, respectively.

INTRODUCTION

Most observational theories assume that PK causes only a small violation of physical laws and thus shows only small effects. Therefore most theoretical parapsychologists do not take RSPK phenomena seriously.

In contrast to the usual observational theories (OT) the Model of Pragmatic Information (MPI) does not start at the description level of quantum theory but at a very general system theoretical level. It roughly says, that psi-phenomena are non-local correlations in psycho-physical systems instead of signals or forces. Such non-local correlations, however, limit the psi-effects due to the conditions of the psycho-physical system, which mainly described by the "meaning" of the situation i.e. pragmatic information (see Lucadou 1995a, Lucadou 1995b).

We have shown in a previous paper that the model of pragmatic information (MPI) at least does not limit the effect size of PK effects even though the model does not assume strong violations of physical laws and moreover assumes that PK is not a real force but a non local correlation. However, the effect size depends in this model on the "dimensionality" of the time order of the PK-events (Lucadou, 2000). This means that the structure and duration of an RSPK-case (its history) and the number of hierarchical levels, which are necessary to describe the organizationally closed system, determine the size of the emerging psi-effects. In an RSPK case, however, one cannot say much about this construct except that all RSPK cases have their own history and time development thus it can be assumed that the dimensionality will be rather high in comparison with PK experiments. In this paper, we will argue that the model of pragmatic information allows one to predict the temporal development of RSPK phenomena and to give instruction to the persons involved in order to cope with the phenomena. In the practice of the "Parapsychological Counseling Office at Freiburg" this model has been applied in dozens of cases with success, however, a detailed statistical evaluation would be an aim of future research.

It is not the aim of this paper to give an overview of the relevant literature concerning RSPK-research, this can be found in (Lucadou, 1984) and more recently in (Horan & Lange, 2000) or with qualification in (Roll, 2003) [1]. However, it must be annotated that in the American literature of RSPK-research the system-theoretic approach to RSPK is nearly completely ignored, in spite of the fact that most of the relevant literature has been published in English speaking (refereed) journals (see for instance: Lucadou, 1983, 1984, 1989, 1991, 1995a,b, 1998, Lucadou & Kornwachs 1980, 1983).

Here, we will first discuss the problem of elusiveness and the dynamics of RSPK-phenomena, before we develop the theoretical tools to describe it. In order to show that these theoretical tools are really able to allow predictions, which can be used to help persons involved in RSPK-cases, we will discuss a single case (of a sitter-group experiment) were the difference between our systemic approach to the usual psychological one becomes evident. Finally we describe two types of RSPK-Cases. From all this we can conclude, that the MPI is not only able to cover the whole range of psi-phenomena, but additionally describes their dynamics and their confusing elusiveness in a coherent way.

ELUSIVENESS

Anyone who takes reports on spontaneous psychokinesis seriously can hardly consider these phenomena to be random fluctuations. The observed effects seem to be too massive and "evident". Is it only chance if a bookcase of over 175 kilograms moves itself 30 centimeters, as was reported in the Rosenheim case? Or if stones as large as a hand fly through the air, as in a poltergeist case in Frankfurt?

Nonetheless, we should bear the following in mind: imagine that someone came up with a conventional color television at the time electromagnetism was discovered. Let us further imagine that someone asked how this apparatus worked and that he was given the answer that it could be explained by the effect discovered by Faraday. Probably Faraday himself would have found it a ridiculous answer, since he assumed that the electromagnetic effect he had discovered was far too weak to serve any future practical purpose.

These facts can serve as a foundation for a rule, nicely formulated by, among others, German parapsychologist Ulrich Timm (1981):

> Reports on extremely strong or stable psi performances contradict general experience and very likely allow for the conclusion that at least part of these are based on conjuring, errors and fraud. We do not claim, however, that such exceptional psi results are in principle impossible. After all, a rule of experience only leads to conclusions as to how likely an effect is. These conclusions can, however, be wrong in specific cases. On the other hand, the rule of sticking to critical scientific rationality may be very useful in an area constantly threatening this rationality. (p. 207)

In practice, "Timm's rule" has proven very successful. Many a time and oft the literature of parapsychology has reported spectacular cases of psi, but after thorough investigation only very few hard facts remain (see Parker, 2002). As such, these facts may still represent a very interesting scientific anomaly, but they prove to be far less spectacular than they seem at first glance. Poltergeist cases are the best example. Quite often these cases evoke the impression that they withdraw from investigation, the "elusiveness". No matter how many unanswered questions and points of critique the Model of Pragmatic Information is confronted with, its adequacy is supported by the fact that it can, at least qualitatively, describe the strange elusiveness of the phenomena rather well. Let us now try to apply this model to poltergeist phenomena.

If an event is rare, it is by definition difficult to observe, for the simple reason that one cannot prepare to do so, a feature that seems to be quite pronounced in poltergeist cases. Hans Bender, the father of poltergeist research in Germany, wrote: "Efforts to photograph or film on-going poltergeist phenomena, or

[1] It seems that W. G. Roll is "actively" ignoring the system-theoretical approach. He cites (e.g. in Horan & Lange 2000) nearly all relevant German literature, but with the remarkable exception of all relevant system-theoretical papers.

to record noises on audio tape are hampered by the problem that these phenomena seem to withdraw from critical observation. One can hardly escape the impression that the intelligent forces hoodwink the observer, by producing a materialization exactly at a spot which cannot be recorded or photographed." (Bender, 1952, p. 169)

Elsewhere, Bender (1974) describes a typical case:

> We were told that firemen ... were posted as guards in the house of K. One of them had observed that at the very moment he turned his head, a pool of water appeared on the floor of the kitchen. (p. 139)

Many researchers have reported such observations, however, they seem to be underrepresented in literature. For instance, in the analysis of about 500 cases, Gauld & Cornell (1978) do not even mention the elusive character of RSPK-phenomena. On the other hand there exists not a single reliable photographic or video document of an "active" poltergeist (Parker, 2002). A plausible explanation for this inconsistency would be that, the elusiveness is not regarded as a property of the phenomena but as an insufficiency of the documentation, or an insufficiency of the possibilities of the researcher, or in general a psychological problem, which has nothing to do with the physical RSPK-phenomena. This view is typical for the physicalistic interpretation of psi, which starts from the assumption that the psychological and the physical part of psi can be separated [2].

THE DYNAMICS OF RSPK

Apart from the elusiveness of poltergeist phenomena, there is a series of other repeating patterns to be found in the literature on parapsychological phenomena (Huesmann & Schriever, 1989). Essentially, they refer to inexplicable noises, objects moving in strange patterns, objects disappearing out of and returning into closed rooms or containers, etc. Many of these events seem to be related in space and time to a youngster in his or her teens, the "poltergeist agent" or "focus person". The events also demonstrate specific time patterns. Generally, their onset is completely unexpected and they develop dramatically. As long as those involved believe that the events are due to external factors, like someone who is fooling them, impulses in electrical circuitry, leaking pipes, etc., the phenomena become stronger and grow into a real demonstration. Those involved feel ever more insecure and try to find external assistance, for example from the police, firemen or from institutions who can provide technical assistance. In this way the phenomena attract wide attention. In many cases there are a number of respectable, reliable and independent witnesses, who feel completely desperate about the causes of the phenomena. We call this the "surprise phase". It is followed by the first hunches that something supernatural might be going on. Indeed, the media, such as newspapers, radio and television show up. Depending upon the socio-cultural background, the phenomena may be attributed to phantoms, spirits, the deceased, witches, poltergeists and parapsychological powers. Only at this point do parapsychologists have the opportunity to get involved. Quite often the previous phase of hunches has already attributed the phenomena to one or more persons, and has coupled general desperation and anxieties with curiosity: the "displacement phase". During this phase, the interpretation of the phenomena shifts from external to internal sources. The same displacement takes place in the phenomena themselves. New types of events manifest, replacing those that had become familiar. As bad as matters are, worse is still to come. Journalists hungry for sensation, self-appointed "parapsychologists" or "exorcists" will plague those involved. To the external curiosity is added an ever-stronger pressure to reproduce the phenomena, which are still strongly confirmed by the initial eyewitnesses. The stronger this pressure grows, sometimes even enhanced by the parapsychologists who rush to the scene, the less the phenomena occur: the "decline phase" has begun. Many of those who expected sensational effects are now disappointed and leave. Often enough, the person who evoked the events is found to make use of manipulations or fraud during this phase. In personal communication, Bender noted that it was his

[2] Even though the problem of elusiveness came to the attention of some psychical researchers, it was not regarded as a hint to the systemic nature of psi. This can be seen in Kennedy's work (see Kennedy 2001), who included this aspect only after Bierman (2002) made a plea of it (see Kennedy 2003).

experience that in almost all cases this phase leads to manipulation, or that this can at least no longer be excluded, especially because the phenomena occur only seldom, in confusing situations or in the direct vicinity of the evoking person. Decline is followed by the final phase of poltergeist cases: "suppression". Fraud is more or less openly discussed, the people and witnesses involved are often ridiculed and discriminated in the mass media, witnesses may even deny (in court) their previous statements and debunking articles are published. The process of social suppression starts: a "conspiracy to cover it up", as Fanny Moser (1977, p. 30) termed it.

In an attempt to describe the psychological aspects of poltergeist phenomena, it is extremely important, as Bender emphasized time and again, to consider the possibility that someone is playing tricks. Its character is aggressive, regressive, and often atavistic. In this context, the psychodiagnostic investigation of the agent is of interest. Freiburg psychologist Johannes Mischo (1970) reports a number of common characteristics: actual conflicts, psychological lability, short but strong irritability and a low level of tolerance for frustrating events. Bender regularly pointed out that poltergeist phenomena should be understood as an unconscious plea for help. The main advantage of a systems-theoretical approach is that it is not necessary to check the reality of each phenomenon immediately, because the essential interactions between those involved in the phenomena and their observers need not be paranormal. It is useless to organize a hunt after each stone flying around, in order to calculate its trajectory, for example. Far more important is the meaning of the event, which can incorporate normal as well as paranormal effects.

THE CONSTRUCTION OF A SYNTHETIC RSPK-CASE

Figure 1 represents a simplified model for poltergeist cases, a hierarchically interconnected system of organizationally closed subsystems. Each of these exchanges pragmatic (= meaningful) information with its enclosed subsystem.

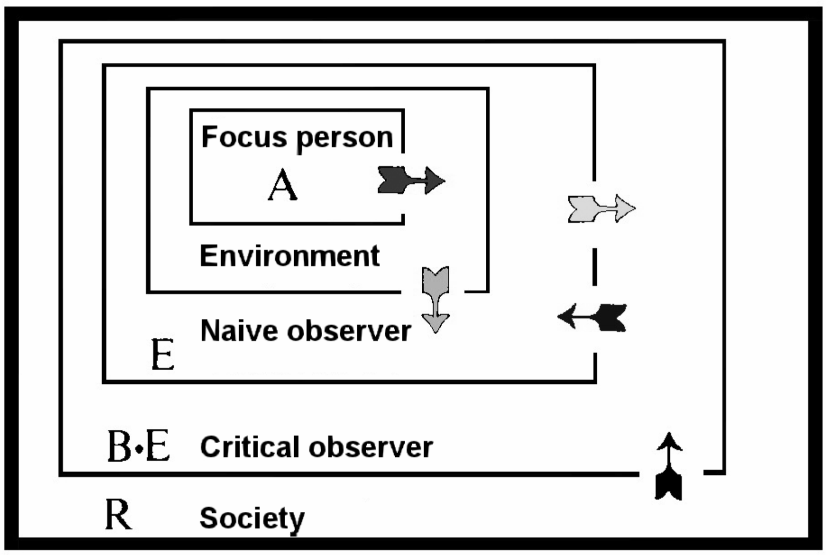


Fig. 1 Hierarchical poltergeist model

The different hierarchically nested subsystems describe the path of the RSPK-action which is exerted by the focus person to different observers and finally to the society. Its pragmatic information, however, flows in both directions, because each observer and the society are also "preparing" the observed system. This can also be described in a formal way: Equation (1) can be considered as the fundamental description of the RSPK system[3].

$$\mathbf{R} * \mathbf{A} = \mathbf{B} * \mathbf{E} = \mathbf{B}' * \mathbf{E}' < \mathbf{I} \quad (1)$$

R is a measure for the "reliability" of the observed phenomena;
A is a measure for the "autonomy" of the RSPK-system;
B is a measure for the "confirmation" (Bestätigung), which is given by an observation;
E is a measure for the "novelty" (Erstmaligkeit) of an observation.

This fundamental equation describes the action of given pragmatic information on a system. Since **R** and **A** or **B** and **E** are corresponding pairs of complementary observables nothing can be said on a partitioning of an outside pragmatic information **I** to reliability **R** and autonomy **A** "inside" of the system without further specifying the "measurement".

From equation (1) one can conclude that any piece of pragmatic information **I** which interacts with a system also produces pragmatic information, but with a new partitioning of novelty **E'** and confirmation **B'** which can be interpreted as the "reaction" of the (sub-) system. Therefore, it is an important question to specify the subsystem – or to be more precise – the boundary of the subsystem (for details see Lucadou, 1983, 1995). The four phases of RSK emerge from the dynamics of the partitioning of **E** and **B**. To "construct" the ideal poltergeist by means of this model, we start with the poltergeist agent.

Phase 1 "Surprise"

The agent is struggling with e.g. puberty, a difficult phase in life, and tries to communicate the associated problems to the environment (unconscious cry for help!). The pragmatic information within his message will only have an effect if he succeeds in balancing novelty **E** and confirmation **B** in such a way that he attracts the necessary attention. If we also assume that his nearest environment and relatives remain deaf and blind (if they reacted properly, his cry for help would be unnecessary), the agent may consciously or unconsciously choose to enhance the novelty of his message. He may have experienced that a number of serious requests for help (confirmation) did not work out. As it is generally known that novelty draws attention, it is now badly needed. A subtle trick or a practical joke, as well as a PK effect is very effective in attracting the attention of the environment, as is clearly demonstrated by the surprise phase in the poltergeist case. It is very likely that another continuum exists between "pure" jokes and the "real" poltergeist, used by youngsters of this age who demand attention. The tricks are in no way less important or less harmless, but parapsychologists do generally not investigate them. Whether he fools his environment or uses psychokinesis, the agent can be assured that all eyes will be on him (or the phenomena) during the surprise phase. (See figure 2) But does the environment understand his cry for help? Naive observers search for all possible causes to "explain the unexplained", but they do not recognize its meaning.

A typical example is a poltergeist case in Germany (Lucadou, 1981). Day after day, large numbers of stones flew into a house owned by a family of Italian immigrants. As a result, their home was largely demolished, but no suspect could be found. The father, a tall man, was very afraid and believed that spirits caused the phenomena. When one of the authors tried to gather more information in a police station, he was told that the case had been solved. A police officer had taken position in the hallway of the house, to observe the incoming stones. The three children of the family were in the children's room at the end of the hallway, while the door to the room had been left ajar. The officer sat with his back towards the door. A

[3] For our consideration it is not necessary to give explicit operationalizations of the described variables. In the context of the MPI **R**, **A**, **B**, **E**, and the operators *, < serve as meta-observables or meta-operators which means that they can be operationalized in any system in many ways. The operationalization depends on the specific situation under investigation. (E.g.: **R** and **A** could be measured by the frequency of the observer's attribution of the phenomena to „technical defects"(**R**) or to „ghosts"(**A**)). This is just the advantage of the systemic approach.

stone coming from the children's room had narrowly passed his head. Although he did not see any of the children throw the stone, no one else could have done so. Thus, concluded the police, the poltergeist case had satisfactorily been solved. When the police officer was asked, whether he considered it normal behavior for children to demolish their own house by throwing stones, he stared quite dumbfounded.

Paradoxically enough, the likelihood that the environment finally understands what it is all about is higher in cases of trickery. As soon as the observers find out who is playing tricks on them, there is at least the possibility that one wonders why the "poltergeist agent" acted as he did. In cases of a "real" poltergeist, this possibility is initially neglected, because one tries to locate the sources of the events in a quite different direction, such as technical malfunction, spirits or angry neighbors, meanwhile overlooking the real message. This enables the poltergeist to go on and on, nourished and at the same time disturbed by the attitude of curious misunderstanding in the environment. In this phase, the elusiveness of the poltergeist is already apparent. Of course people want to see events. There are indeed events, but always at a location where one does not expect them to happen, otherwise there would be no novelty. In the end, however, the game of elusiveness and the continuous flow of pragmatic information leads to the understanding that a specific person is "the center of the cyclone".

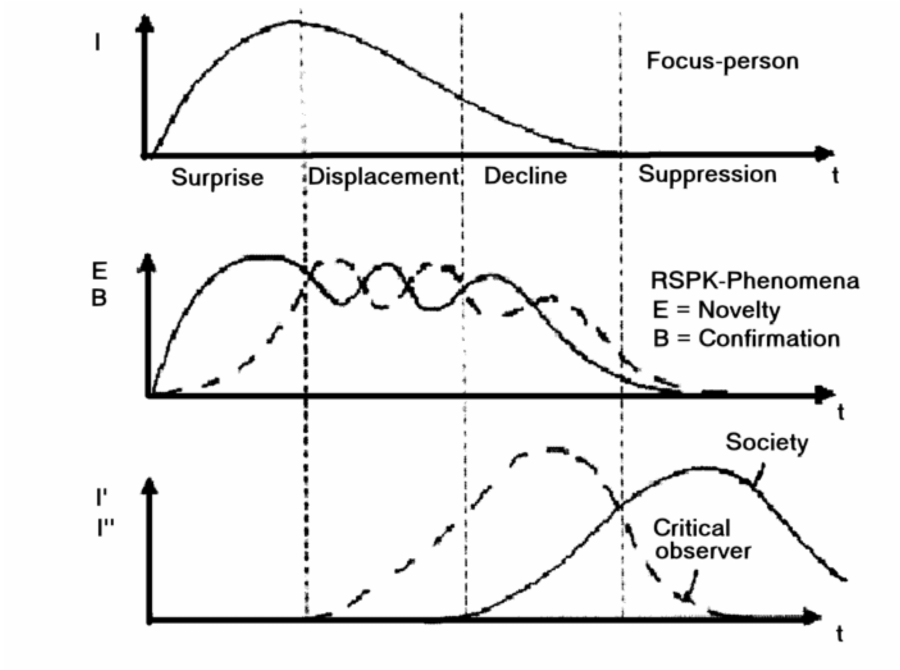


Fig. 2 The four phases of RSPK-activity: The dynamics of RSPK is a result of the complementarity of the autonomy **A** and the reliability **R** of the system, respectively of the novelty **E** and the confirmation **B** of the observations. The subconscious "message" of the focus person shows up in the action on the system, but the "pragmatic information" I is limited. This leads to a continuous change and finally to the disappearance of the RSPK-phenomena (elusiveness). The "critical observers" and the society also exert some action (i.e. pragmatic information **I'** and **I"**) on the RSPK-system in order to "eliminate" the anomaly. The four phases are called: Surprise phase, displacement phase, decline phase, and suppression phase.

Phase 2 "Displacement"

After this Gestalt switch the second phase begins. A beginning is made on a new organizational closure by the naive observers and the person on whom they focus their attention. They join forces and establish one tight group of those who have experienced the inexplicable effects, only to find themselves now confronted with the disbelief and the uncompromising questions of critical observers, such as journalists and parapsychologists. It should be noted here that naive observers need not be uncritical, just as critical observers can be believers in psi. During this phase the focus person, who has in many cases been isolated until then, often becomes the center of attention among the social group to which he addressed his cry for help. In practice, this person is considered special, a medium who has a personal link with the spirit world, a witch or a paranormally gifted person. This would serve his goal, if it were not for the presence of the critical observers in his surroundings, who exert enormous pressure and expect him to produce effects. Suppose that the new organizationally closed unit succeeds in providing these critical observers with pragmatic information. (The message reads: I am something special!) Even in this case it will not generate enough novelty, because the critical observers cannot get enough of it, but according to our model, they also prepare (in the sense of the term "preparation" used in quantum physics) the system because of their observation.

Phase 3 "Decline"

They describe the system because they want to document the phenomena reliably and beyond doubt. To reach this goal, the system needs to be prepared appropriately. Currently, we use video and audio recording apparatus, personality questionnaires and Rorschach tests to do so. Is it then any wonder that the autonomy of the system disappears and that it can no longer exert any scrap of novelty that could be interpreted as psi? In other words: If $\mathbf{B}$ is increased $\mathbf{E}$ must decline since $\mathbf{E}^*\mathbf{B} < \mathbf{I}$ and $\mathbf{I}$ is limited.

Nevertheless, the critical observers are still interested in the phenomenon (often less in the persons involved). One cannot exclude therefore, that these observers too still manage to experience something of the original novelty. It is more likely that something has been made up for them, though certainly not with negative intentions. The result can even represent a joint venture of the group as a whole. The remarkably simple moral justification for such manipulations is often the fact that those involved know what has "really" happened and do not hesitate to lend the phenomena their support, in order to enhance them a little. Among the members of the group, the neat distinction a scientist makes between simulation and reality may not be that important. For this reason the parapsychologist who uses the in itself very acceptable method of participant observation, should be very careful.

In many cases, however, it is too late to do so during the decline phase, since another and permanent level of the system has come into being. It represents the border between, on the one hand, the organizationally closed system of those who are interested in the poltergeist phenomenon and, on the other hand, society in general.

Phase 4 "Suppression"

The phase of suppression and denial has started. Neither society nor governmental institutions are fond of the anarchy of poltergeist cases. Their objective is to command (or govern) reliable systems. A poltergeist is indeed a very unimportant example of the "preparation" effect of society. Social psychology calls this effect "social perception". Experiments show that pressure from society can alter the way in which individual members of the society perceive a particular phenomenon. One can drive someone to "see a U where there is an X". It is, after all, no coincidence that (at least in Germany) jurists, people in forensic medicine and police officers, etc., feel the need to eradicate parapsychology root and branch.

Duration and structure of RSPK in the light of the MPI

According to the Model of Pragmatic Information, the components of pragmatic information in poltergeist cases are alternatively determined by two goals. The first goal is the internal goal of the organizationally closed system to produce an effect in society. The second one is the external goal of society to prepare the organizationally closed system. The internal goal affects the combination of novelty and confirmation, while the external preparation of the system has an effect on the autonomy and reliability of the system. Only autonomous systems can produce novelty, a necessary component for a phenomenon to be evaluated as anomalous or psychical. To preserve the autonomy of the system, however, one should not prepare it in such a way that everything has been determined. The system "can only behave as it pleases" as long as one does not observe it with great care. A predetermined system loses its autonomy and because of that it loses its ability to be unique as well. For this reason, a poltergeist, which can reliably produce "recurrent spontaneous psychokinesis" (RSPK) for an extended period, needs to be very "strong" indeed. This conclusion neatly fits the experimentally supported "rule of Timm".

The structure of RSPK is governed by a very fundamental assumption of the MPI which suggests that the complementarity of autonomy and reliability originates from an even more fundamental complementarity, that of structure and function. We could even say that all things we are able to describe are described according to these categories. It depends upon the author of the system which of the two aspects is emphasized.

As an example, let us use the description of a radio receiver. For the user, a general description of its functions will do. He only wants to know what purpose the apparatus serves, how to switch it on, how to find a particular station and so on. He is hardly interested in the structure of the receiver. As long as it works, he does not care about the internal construction of the equipment. In case of malfunction, however, the user may suddenly begin to show interest in the internal structure. The repairman cannot do without the description of the electronics in the receiver, a structural description with the greatest possible accuracy, to be able to restore the proper functioning of the apparatus. This description as such is not sufficient to trace the fault. The repairman also needs to know what changes in the machine's behavior were observed, how it malfunctions. The owner should tell the technician which functions do not work properly. Only after the two categories of description match the system has been described in an optimum way. Probably each of us has brought a defective piece of equipment to a repairman: when the technician takes it apart and reassembles it, no defect can be traced, and no one can say why. In general, a change in the functioning of the system, a malfunction, is experienced as something unique. Suddenly, the apparatus "does not work as it ought to". If, however, we can attribute the changed functioning to a changed structure of the system, after a while we find confirmation, because the new "system configuration" no longer changes. Suppose that a tube in a receiver has blown. In this case the structure of the system has changed and it produces only confirmation: it does not work any more. This proves that functional and structural changes are linked to the concepts of novelty and confirmation and therefore to autonomy and reliability as well. One of the basic statements of systems theory holds that these categories of description do not simply represent properties of the system to be described (its perception of the world), but that they represent the system itself.

In poltergeist cases, we can test this hypothesis empirically. If structure and function were merely categories of the observer, we would expect to find these categories in different quantities in witness reports on poltergeist cases. If we combine reports of different witnesses (statistically), the unimportant differences of the reports should vanish. In the long run, therefore, the characteristics of the poltergeist itself will become visible. (The same thing happens when a police officer tries to reconstruct the causes of an accident by interviewing different witnesses, each of whom has his own point of view.)

A recent analysis of witness reports on poltergeist cases actually demonstrated that these reports contain two factors, which can be labeled "structure" and "behavior". These factors even persist if one takes the mean of several observers of a poltergeist case. This result teaches us that structure and function are really two categories of the poltergeist itself, instead of merely belonging to its observers. This hypothesis of the Model of Pragmatic Information was stated before the experiment started, and the two psychologists who carried out the research, Monika Huesmann and Friederike Schriever (1989), were not aware of it. For this reason they cannot have adapted the interpretation of their data to make them fit the hypothesis. Furthermore, they developed an objective procedure in which all witness reports of over 60 cases were split into parts (so-called items) and then analyzed by a statistical procedure (factor and cluster analysis). This analysis led to the two basic factors discussed above, which were predicted by the model. From this, we can conclude that the Model of Pragmatic Information is also capable of predicting aspects of "macro PK phenomena", though these conclusions have more of a qualitative than a quantitative nature. It teaches us more about the limits or restrictions of the phenomena, than about their strength.

IS ELUSIVENESS A PSYCHOLOGICAL OR SYSTEMIC EFFECT?

It seems logical to extend the findings of the MPI across the limits of experimental psi effect or poltergeist phenomena, into other situations dealing with qualitative macro-psychokinesis. In so doing, we find that the model need not even be afraid of the darkened rooms in the era of physical mediums. Moreover, we can find that another prediction of the MPI can be tested empirically, namely a kind of "uncertainty relation" between the "effect-size of the phenomena" and the "quality of their documentation":

Effect-Size of a psi Phenomenon * Quality Of its Documentation < I (2)

According to the MPI this equation holds for all types (experimental, RSPK, sitter group) of genuine psi effects. However, in general it is very difficult to distinguish between its common psychological interpretation and the systemic interpretation of the MPI. Fortunately, in a single case of a sitter-group experiment this important distinction could be demonstrated. (It goes without saying that further research is required. However, in the practice of the Parapsychological Counseling Office in Freiburg we have used this theoretical model in dozens of cases with remarkable success. [4])

The English psychologist Ken Batcheldor (1979) developed a psychological model, which might explain why the phenomena observed by a group of sitters are so much stronger than the results of experiments with single subjects. Batcheldor's model suggests two inhibiting factors, which generally stop people from exerting their psychokinetic abilities. The first of these is ownership resistance, the fear of possessing psychokinetic powers. The second factor is witness inhibition, the fear of witnessing a paranormal phenomenon. Batcheldor assumed that even if people are not aware of these fears, they can even play a role, which means that if someone is convinced he is not bothered by any of them they may be subconsciously active. Even parapsychologists suffer from these fears, let alone skeptics. It seems quite obvious that such a psychoanalytically oriented characteristic leads to the danger of self-immunization (a non-falsifiable hypothesis), particularly if it is used to explain the absence of an effect.

The main aim of Batcheldor's model was to investigate under what circumstances both defense mechanisms could be switched off. According to his numerous experiences one of the best conditions is the technique of the sitter group, since it resembles the classical mediumistic séance, though often without a professional medium and not necessarily using the ideology of spiritualism. According to Batcheldor, this technique relieves each of the sitters from the burden of personal responsibility for the phenomena. Each sitter can always feel himself a relatively uninvolved observer. In other words, it is not he, but the other sitters who produce the phenomena. Furthermore, the possibilities of observing effects in a group are in no way as good as many tend to believe. After all, the sitters are expected to create a relaxed and jolly atmosphere. This is especially so for séances in darkness or under red light. In many cases a trickster is elected beforehand, to simulate "phenomena" and trigger off the real thing. Needless to say, this sounds somewhat suspicious. On the other hand, it can help to reduce witness inhibition among other participants, because they do not know for certain whether they are seeing only a trick or a paranormal phenomenon. It will be clear that the experimental setup should nevertheless offer the opportunity of distinguishing tricks from real phenomena. From a psychological point of view, it is very interesting to note that spiritualistic ideology, too, can contribute to the elimination of psychological defense mechanisms. If the sitters are convinced that spirits of the deceased produce the phenomena, they do not feel responsible for these phenomena and their origin. This resembles spontaneous paranormal cases, where the alleged spirits and demons can serve to suppress the real problems.

[4] The aim of the therapeutic approach is of course to reveal the unconscious problem of the focus-person, however, sometimes an immediate intervention is so urgent, that we recommend to the persons concerned to install a video camera, which is able to "observe" the whole room and the camera itself in a mirror. With such an installation the RSPK-phenomena immediately vanish at least in this room – to the incredulous astonishment of the persons involved (see below).

One can even combine Batcheldor's technique of the sitter group with an experimentally produced spiritualistic tale, to yield fantastic results. A group around the Canadian parapsychologist Iris Owen did such an experiment, in which the sitters deliberately "constructed" an artificial spirit. To do so, they invented the tragic-romantic story of an English squire by the name of Philip, who lived in the middle of the 17th century. The details of the story were carefully elaborated. At the same time, however, the group took great care to prevent a resemblance to any existing historical entity. The fictitious squire Philip fell in love with the beautiful gypsy girl Margo, but could not face the consequences of this romance, and finally committed suicide. Since then, he has been a spirit, and was conjured up during the Canadian pseudo-spiritualistic séance. And indeed, he came into being, produced rapping sounds in the table and finally demonstrated the whole spectrum of spiritualistic manifestations. The complete experiment has been described in an interesting book "Conjuring up Philip" by Iris Owen and Margaret Sparrow (1976). The book clearly demonstrates that spiritualistic phenomena can also be produced by invented "pseudo-spirits". In this respect, it is an impressive illustration of Batcheldor's model.

On the other hand, there are observations, which do not readily fit into Batcheldor's model, which holds mainly psychological factors responsible for the elusiveness of the phenomena in the era of physical mediums. If only psychological factors could be held responsible for the avoidance of a paranormal phenomenon, it would suffice to eliminate these factors. Indeed, many parapsychologists argue that the elusiveness of the phenomena is only an "irritating by-product" of unfavorable psychological conditions, which should be eliminated. This is why authors of reports often simply neglect the remarkable elusiveness, upon which the astonished reader wonders why such massive phenomena were not "properly" investigated with all the technical facilities available to the experimental sciences. Not until such a reader requires more precise answers, is he informed that the phenomena behave in such an elusive way that they can hardly ever be properly objectified. Even after many years of effort, Batcheldor could not lay his hand on a single video recording of a paranormal phenomenon. In a letter to one of the authors, Batcheldor (1985) described a typical situation:

> During an experiment we had switched on the infrared video camera when the table levitated. Although we thought the video recorder was running, we did not feel inhibited and I believed we had achieved a success. When we played back the tape, however, it did not contain any images at all! We found out that a switch had been in a wrong position. The next time I carefully checked the position of all switches and, indeed, the table refused to levitate. So would it be possible, though difficult according to your theory, to acquire a detailed video recording of a levitation?

According to the Model of Pragmatic Information, we should indeed assume that the psi effect remains elusive even when the psychological barricades have been eliminated, because every observation seeking confirmation prepares the system in such a way that its autonomy is restricted. In the situation described by Batcheldor, the psychological conditions were in fact favorable and levitation was observed. Nobody knew about an erroneously set switch and it is difficult to understand why after the correction of this mishap the psychological situation would have changed so dramatically. According to our model, however, the position of that switch, which is of no importance for the psychological situation, is of fundamental importance. Because of the erroneous setting of the switch the total system was objectively unable to make a recording of the phenomenon. In other words, any measurement or recording was impossible. The phenomenon could only occur because it was not completely objectifiable. A complete video recording would comprise more pragmatic information than the system was able to produce. The subjective experience of the sitters, on the other hand, is "diffuse" enough to record the less voluminous or less reliable information of the phenomenon. The situation remains vague, the system is not completely prepared for reliability. Confirmation of the actual phenomena by means of a video recording is lacking. By resetting the switch, the structure of the complete system is altered to such a degree that the phenomenon cannot occur. Switching on an additional measuring apparatus changed the potentiality of the system in such a way that particular (complementary) measurements were prevented from taking place. It goes without saying that the above does not mean that the Model of Pragmatic Information considers any objective observation of psi phenomena absolutely impossible. This would represent an inadmissible immunization. One could summarize it as follows: less equipment may have produced more phenomena! It is necessary to adapt the objective conditions of observation to the phenomenon, in such a way that enables the observer to gather the optimum pragmatic information the system is able to deliver. Without this adaptation, one throws away information. If no phenomenon occurs when a complete video recording is made, one apparently threw away too much information about the phenomenon: it does not occur any longer. If, on the other hand, a séance in darkness prevents the observer from separating trickery and real phenomena, then he has also thrown away too much information, because he does not know what he has observed. Where is the royal road between the Scylla of an observation and the Carbides of a phenomenon without observation?

In answer to Batcheldor's letter, it was suggested to reduce the resolution of the method of observation, i.e., to defocus the video camera or to limit it to documentation on audiotape only. In this sense, "less" would really be "more", because we would obtain an objective recording of the phenomenon, which would be less easy to interpret than a perfect documentation, because it contains lacunae. These are exactly the ambiguities resulting from an imperfect method of recording. Suppose that only noises are recorded, in which case the causes of the noise remain unclear. If a camera is out of focus, it reduces the possibility of determining the exact location of any phenomenon. This is not to say that lacunae in the documentation procedure should leave room for manipulations or fraud (for example because one can no longer see any wires used for trickery). They should rather prevent the system from being prepared too unilaterally for reliability, so that it loses its autonomy. This method of recording reduces the pragmatic information and offers the experimenter fewer opportunities to utilize the phenomena in the sense of signal transfer. He has only a limited degree of control over the system. As the recording does not teach him exactly what is going on in the system, he cannot undertake any goal directed actions. In one way we are familiar with this connection since the era of physical mediums, but up to now it was interpreted in a completely different way. The pioneers of that era had already found out that the phenomena "feared daylight" and were only produced in darkness or under red light. From their point of view, however, this could be attributed to the danger light posed to a "fragile" substance, the existence of which was a necessary condition for the phenomena to occur. This substance would be destroyed by daylight. Batcheldor, too, reported that he could often make an audio recording of his experiments, but that no phenomena occurred when video recordings were made.

In the experiments Batcheldor performed till shortly before his death, he tried to utilize observations with differing degrees of resolution. He reported darkened sessions as well as the use of a fluorescent background panel, in front of which "cloths could materialize in the air". As predicted by our hypothesis, in front of a panel with a grid of fluorescent dots these materialized objects (or whatever they were) remained visible for a longer period than in front of a panel completely covered by fluorescent paint. In rare cases (infrared) video recordings succeeded. In these cases levitated objects were always in such a position that it was impossible to decide whether they were really levitated or only held in front of the camera. They seemed to have been put in a position that prevented the observer from finding out how the phenomenon came about. Batcheldor emphasizes that it would have been very difficult for the sitters to manipulate the object in this specific position, as they did not know what visual field was covered by the camera, which was not equipped with a viewfinder. These manipulators could, therefore, easily have been detected. Batcheldor's impression is that the complete system "knew" exactly what was recorded by the camera and that one could only record a phenomenon if its cause remained hidden in darkness, so that it was impossible to decide whether a normal or a "paranormal" event had taken place. This is exactly the same as not being able to interpret the video recording. It contains less pragmatic information and prevents the experimenter from having complete control over the system or from making it reliable.

The discussion of Batcheldor's approach in the light of the MPI shows that the notion of (more or less complete) "control" is an important issue for the structure and limitations of PK-Phenomena. Again, this feature also shows up in RSPK-cases.

TWO TYPES OF EXPECTED RSPK PHENOMENA

As we have argued above RSPK phenomena are mainly a subconscious action of the focus person, which can be compared with a psychosomatic reaction in his or her environment. However, we found in the parapsychological counseling project some cases, which do not fit with this assumption. (It will be the aim of future research to collect such cases systematically and to provide a statistical analysis; here we are discussing this issue for the reason of completeness of the MPI).

Some persons suffer from RSPK activities, but no subconscious problem seems to be present. This is, however, not a hard criterion because one can never guarantee whether this assumption is justified. But the model can easily be extended to such cases.

In these cases the focus person seems to be much more passive, for instance, it is very often a person who suffers from depression and is not able to control anything in his or her life and also not in her environment. This in contrast to the active RSPK focus person where one gets the impression of a boiling pot, which is ready to explode, and the phenomena are just the sign of an "explosion". With the passive ones the opposite is the case.

This feature can be included in our model by using the concept of "system control". System control describes how an organizational closed system controls itself by interaction with the environment. In the active case the focus person serves as the "master" part of the control-cycle and the environment as a "slave". In the passive case it is just the other way round. The focus person is not able to control anything and is not able any more to stabilize his or her world. The active focus person in contrast is even over-controlling his or her environment, which leads to macroscopic random fluctuations i.e. RSPK phenomena.

We assume here that everybody under normal circumstances subconsciously controls his or her environment to stabilize it. This means that random fluctuations, which are too large, are suppressed. This can be seen, for instance, in the sitter-group experiments discussed above and in many PK experiments with subjects who do not get deviations from chance but instead a smaller variance. This means that natural systems themselves – whatever this means – may produce larger fluctuations if they are not observed. (In quantum physics this is known as "Quantum-Zeno-effect": "An observed pot never boils"). It is a fundamental assumption of the model of pragmatic information that observation and also non-observation are different preparations of the system. This idea can also be found in folkloristic reports that spooky events seem to happen at unobserved places, for instance, that houses where nobody lives crumble more rapidly than if they are in use.

From this point of view it could be assumed that in a case of depression the person loses more and more the control on the organizational closed environment, which shows up in fluctuations within this environment. This means that the integrative power, which keep the whole system together, cannot suppress individual fluctuations within the system with may destroy the system. In this case it is expected that this will occur only in a certain period before the whole system dies out. In such cases we have no real displacement phase but only a decline phase and even the suppression phase is not necessary. The decline phase is not driven by the attempt to produce the phenomena but simply by exhaustion.

Therefore the active RSPK case can be considered as an immune reaction of the whole system caused by an unconscious problem, which acts like a "virus", where as the passive RSPK case must be considered as a disease, which leads to disintegration of the system.

In clinical terms, the active RSPK can be considered as a phenomenon of dissociation[5] whilst the passive RSPK phenomena show the phenomena of depression and degeneration. This fits exactly with the assumption of the model of the pragmatic information that psi phenomena correlate with the temporal change of complexity and not with complexity itself.

[5] In the Parapsychological Counseling Office at Freiburg, over years, we have indeed collected strong empirical evidence, that RSPK-focus-persons show a high degree of personality dissociation (disorder).

CONCLUSIONS

The examples discussed above show that there is not too much difference between the phenomena produced by physical mediums and by poltergeist phenomena if we consider them within the MPI. Both types of phenomena are difficult to observe and they are dependent upon the preparation of the system. There are also common aspects in their psychological description. In both cases there is the dynamical process within a group, in which the interaction among its members is of fundamental importance, though often inadvertently in poltergeist cases.

It will be clear that we are still far from being able to say which phenomena in the area of macro-PK are possible or impossible, or to conclude why they are as they are and not different. We simply do not have a clue how strong the "stochastic fluctuations" of a system can become and why they often seem so bizarre. After all, such a group represents a system with an immense degree of entanglement and it comprises many levels of description. Even if properly controlled PK experiments with random generators yield so small a result that it is hardly detectable, it need not be small in real life too. A controlled experiment represents a very artificial, more or less sterile situation, offering no opportunity of drawing direct conclusions as to what occurs in daily life. In the EPR experiment, too, the difference between the (invalid) classical theory and the (correct) quantum mechanical description can only be detected by means of complicated measuring techniques in a most artificial situation. The concrete effects of quantum physics, however, do not only steer our modern technology, but all the phenomena in nature, which we cannot describe if classical physics were universally valid. Once we have accepted that the descriptive language of psychology, too, is of a "non-classical nature", then many at first sight simple events suddenly look quite different. It seems to me, that we are still light years away from real understanding of the mysteries behind poltergeists and physical mediums.

Nevertheless, the Model of Pragmatic Information seems to provide some concepts with which we can formulate experimental hypotheses and which can in principle be falsified. Whenever it could be shown, that psi-phenomena do not suffer from the limitations discussed above (e.g. a deliberate repeatable PK-action) the model would clearly be falsified. Furthermore, the MPI does not only enable us to integrate very different areas of parapsychology, but has already made clear-cut predictions, without the necessity of restricting ourselves to empty concepts like psi. These are no more than "empty words", as Hans Bender often used to say. In most cases they rather represent our despair than our understanding.

REFERENCES

Batcheldor, K.J. (1979). PK in sitter groups. *Psychoenergetic Systems 3*, 77-93.

Batcheldor, K.J. (1985). [Unpublished letter to Walter von Lucadou].

Bender, H. (1952). Zur Entwicklung der Parapsychologie von 1930-1950. In: Driesch, H.: *Parapsychologie*. Zürich: Rascher, 3. Auflage, p. 135-176.

Bender, H. (1974). Modern poltergeist research – A plea for an unprejudiced approach. In: Beloff, J. (Ed.): *New directions in parapsychology*. London: Elek Science, p. 122-143.

Bierman, D.J. (2002). [Letter to the editor]. *Journal of Parapschology, 66*, 102-106.

Gauld, A. & Cornell, A.D. (1979). *Poltergeists.* London: Routledge & Kegan Paul.

Huesmann M. & Schriever, F. (1989). Steckbrief des Spuks. Darstellung und Diskussion einer Sammlung von 54 RSPK-Berichten des Freiburger Instituts für Grenzgebiete der Psychologie und Psychohygiene aus den Jahren 1947-1986. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie 31*, 52-107.

Houran, J. & Lange, R. (eds.) (2000). *Hauntings and Poltergeist – Multidisciplinary Perspectives.* Jefferson: McFarland.

Kennedy, J. E. (2001). Why Is Psi So Elusive? A Review and Proposed Model. *Journal of Parapsychology 65*, 219-245.

Kennedy, J. E. (2003). The Capricious, Actively Evasive, Unsustainable Nature of Psi: A summary and Hypotheses. *Journal of Parapsychology 67*, 53-74.

Lucadou, W.v. (1981). A recent alleged RSPK case in Frankfurt. In: Roll, W.G. (Ed.): *Research in Parapsychology 1980.* Metuchen, London: Scarecrow Press, pp. 58-60.

Lucadou, W.v. (1983). Der flüchtige Spuk. In: Bauer, E. & Lucadou, W.v. (Eds.): *Spektrum der Parapsychologie.* Freiburg i.Br.: Aurum Verlag, p. 150-166.

Lucadou, W.v. (1984). What is wrong with the definition of psi? European Journal of Parapsychology. 5, 261-283.

Lucadou, W.v. (1989). Vom Abgrund der Systeme. Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie 31, 108-121.

Lucadou, W.v. (1991). Some Remarks on the Problem of Repeatability of Experiments Dealing with Complex Systems, in: Atmanspacher H. & Scheingraber, H. (eds.) Information Dynamics. NATO ASI Series, Series B: Physics Vol. 256. New York & London: Plenum Press. pp. 143-151.

Lucadou, W.v. (1995a). The Model of Pragmatic Information (MPI). *European Journal of Parapsychology 11*, 58-75

Lucadou, W.v. (1995b). *Psyche und Chaos – Theorien der Parapsychologie*, Inselverlag, Frankfurt a.M.

Lucadou, W.v. (1998). The Exo-Endo-Perspective of Non-locality in Psycho-Physical Systems in: Casys, International Journal of Computing Anticipatory Systems 2, 169-185.

Lucadou, W.v. (2000). Backward Causation and the Hausdorff-Dimension of Singular Events, in: Steinkamp, F., (ed.), *Proceedings of Presented Papers*, The Parapsychological Association 43rd Annual Convention August 17-20, 2000, Freiburg i.Br. p. 138-147.

Lucadou, W.v., Kornwachs, K. (1980). Development of the Systemtheoretic Approach to Psychokinesis. European Journal of Parapsychology 3, Nr.3, 297-314.

Lucadou, W.v., Kornwachs, K. (1983). On the Limitations of Psi. Psychoenergetics 5, 53-72 (Abstract in: W. G. Roll et.al. (eds.): Research in Parapsychology 1982. Metuchen, N.J. & London (1983) 85-89).

Mischo, J. (1970). Zur Persönlichkeitsstruktur psychokinetischer Medien. *Zeitschrift für Parapsychologie 12*, 19-25.

Moser, F. (1977). Spuk – Ein Rätsel der Menschheit. Olten u. Freiburg, Walter-Verlag.

Owen, I.M. & Sparrow, M. (1976). *Conjuring up Philip.* Ontario, Canada: Fitzhenry and Whiteside.

Parker, A.. (2002). Cognitive Psychology's Day in Court: An Essay Review of Houran and Langr's Haunting and Poltergeist – Multidisciplinary Perspectives. *European Journal of Parapsychology, 17*, 97-108.

Roll, W.G. (2003). Poltergeists, Electromagnetism and Consciousness. *Journal of scientific Exploration 17*, 75-86.

Timm, U. (1981). Die Unwahrscheinlichkeitsregel – oder: Können Psi-Resultate „zu gut" sein? *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie 23*, 207-217.

Address for correspondence: Walter von Lucadou,
Parapsychologische Beratungsstelle der WGFP,
Hildastr. 64, D-79102 Freiburg i.Br., Germany.
E-mail: info@parapsychologische-beratungsstelle.de

ANTICIPATORY SKIN CONDUCTANCE RESPONSES: A POSSIBLE EXAMPLE OF DECISION AUGMENTATION THEORY

Edwin C. May
Laboratories for Fundamental Research
Palo Alto, California USA

ABSTRACT

Two recent skin conductance response studies have provided significant evidence of an effect. In both studies, the dependent variable was a proportion difference between prestimulus intervals prior to acoustic stimuli and prior to controls that contained a fully-formed non-specific skin conductance response (ns-SCR). The null hypothesis was that the proportion difference should be zero. Both studies showed significant non-zero proportion differences. These studies have been examined with regard to two competing hypotheses for a possible mechanism. A physiological hypothesis suggests that individuals respond in advance to randomly-determined future startle stimuli. A second hypothesis suggests that the participants are simply a source of random ns-SCR's that are not due to any overt stimuli and that the observed effects arises because of the experimenters' psi. That is, Decision Augmentation Theory (DAT) holds, in this context, that the experimenters bias their decisions to initiate an experimental run such that the proportion difference will be large and thus mimic a physiological response. Circumstantial evidence against a physiology hypothesis include: the rare anticipatory response to either an acoustic stimulus or a control is independent of stimulus position; previously certified laboratory psi participants did not perform better then unselected participants; and participants who show no or very few ns-SCR's (i.e., zombies) do not show anticipatory effects at all. Quantitative analyses do favor a DAT hypothesis. DAT predicts and the data significantly confirm that the ns-SCR rate prior to formal controls will be less than that measured far from any stimuli (i.e., background rate). Additionally, DAT predicts and the data confirm a square root dependency of Z-score versus lability (i.e., the probability of a ns-SCR in a 3.5-second period). The circumstantial arguments taken together with two different types of quantitative results provide the best possible evidence in support of a DAT mechanism that can be obtained retrospectively. A prospective study, however, is needed settle the issue, formally.

INTRODUCTION

Anomalous anticipatory effects in the human autonomic nervous system were first conducted by Vassy in the late 1960's, but reported later (Vassy, 1978). Vassy used electrodermal activity as a dependent variable in a classical conditioning experiment. The design investigated whether a conditioned response to a stimulus would appear before a randomly timed unconditioned electroshock stimulus. Vassy found that in six of 10 sessions skin conductance prior to the randomly timed electroshock stimuli was significantly higher than before controls ($p = 0.01$). Vassy later successfully replicated this outcome (Vassy, 2003). Hartwell examined CNV's for prestimulus effects in the central nervous system but found no evidence of psi (Hartwell, 1978). More modern studies (Bierman & Radin, 1997; Radin, 1997) independently confirmed Vassy's results using a design based on randomly chosen photographic stimuli with varying emotional content. They found significant differential anticipatory effects in skin conductance levels five seconds before the stimuli were randomly selected and displayed.

May and Spottiswoode improved upon earlier skin conductance experiments by simplifying the design from level shifts and cognitive stimuli in the presentiment work (Bierman & Radin, 1997; Radin, 1997) to Bernoulli counting of anticipatory responses to future startle acoustic stimuli (May & Spottiswoode, 2003; Spottiswoode & May, 2003). In these audio experiments, the dependent variable was the difference in proportions of non-specific skin conductance responses (ns-SCR) observed in 3.5 second regions prior to future acoustic stimuli, as compared to the same measure prior to future controls. Statistical significance was determined by comparing the difference between the two proportions (Utts & Heckard, 2004). May and Spottiswoode reported a Z-score for the proportion difference of 5.08 (n = 1000 acoustic stimuli, effect size/stimulus = z/sqrt(n) = 0.161±0.032). A significant result was also observed in Budapest, HU (May, Paulinyi, & Vassy, 2004) with a conceptually similar design for a proportion difference Z-score of 2.08 (n = 725, effect size/stimulus = 0.077±0.037).

The primary question addressed in this paper is which of the non-chance competing hypotheses represents a better fit to the data.

HYPOTHESIS

Our cherished hypothesis is that the participants in these experiments are reacting in advance to future startle acoustic stimuli. That it, the effect is "force-like" in that a future experience "causes" an autonomic nervous system response in the present.

An alternative hypothesis is based upon an information process—Decision Augmentation Theory (DAT). In 1995, May et al., formally introduced the concept (May, Utts, & Spottiswoode, 1995). This paper was the culmination of years of model building and name changes from Psychoenergetic Data Selection, Intuitive Data Sorting to finally Decision Augmentation Theory (May, Humphrey, & Hubbard, 1980; May, Radin, Hubbard, Humphrey, & Utts, 1985). This model holds that individuals use their psi ability to bias decisions toward more favorable outcomes. This idea was an extension of a similar concept first proposed by Stanford (Stanford & al., 1976; Stanford & Stio, 1976; Stanford, Zenhausern, & Dwyer, 1975).

A reviewer of this paper and Dick Bierman[1] have vigorously adopted a differing view. That is, "no one" any longer considers a force-like interpretation in μ-Pk or in the context of psychophysiology in this paper. This cannot be. If it is claimed that targets of PK activity are physically different than they otherwise would have been as a result of the operator's attention/intention, then forces are responsible. Forces are real even in the quantum domain of atoms. Electrons travel along paths due to E&M forces—otherwise particle accelerators and television sets would not work. If, on the other hand, the claim is that there is some, as yet not understood, correlation between the action of the pk-target device and the attention/intention of the operator, then the device is not different that it would have otherwise been and no force is involved. How such a correlation might exist, is still a mystery; however, DAT attempts to limit the problem space to "simple" precognition.

In the context of anticipatory skin conductance response studies, DAT suggests that the designated participant is simply a source of random skin conductance responses and that psi in the study arises because of a putative DAT skill on the part of the experimenter who initiates each run with a single button press. [2]

THEORETICAL APPROACH

In the formal paper (May et al., 1995), we introduced a test for DAT involving Stouffer's Z:

$$Z_2 = \sqrt{\frac{D_2}{D_1}}\, Z_1,$$

where D_1 and D_2 were the number of potential decision points available to reach an end result of Z_1 and Z_2, respectively. That is, the more decision points there are the more DAT-likely it becomes to obtain a higher Z-score.

[1] Private discussions.

[2] In the acoustic startle experiments, the experimenter presses a button once for a run; differing from most of the presentiment studies, the participant makes no trial or run decisions.

A Monte Carlo technique was used to explore the space of available decision points in a typical acoustic prestimulus response study for a given over all lability (i.e., probability of a ns-SCR). Imagine a single Monte Carlo pass consisting of a run of 30 minutes as a 515-fold array of 3.5-second intervals. Using a Marsagula pseudo random generator (Marsaglia & Zaman, 1987), a specified lability was used to compute which of the 3.5-scecond intervals contained a ns-SCR. Then using the inter-stimulus intervals from the two acoustic prestimulus response studies both the random position and random stimulus type (i.e., acoustic or control) were imposed upon this array.

To measure the observed lability for this pass, consider eight 3.5-second intervals around each of the silent controls—four intervals post control and four intervals prior to the formal prestimulus region. Thus for a single pass with an average of 10 (or 15) controls, this amounts to about 80 (or 120) intervals from which to compute the average. Call this average lability the background rate, BK. Additionally, define P_1 as the ns-SCR rate prior to acoustic stimuli and P_0 as the ns-SCR rate prior to controls. The dependent variable was the ratio of the number of passes that lead to $P_0 < BK$ to the number of passes that lead to $P_1 > BK$. Figure 1 shows the result from 25,000 Monte Carlo passes as a function of lability.

The solid dark line represents the ratio for ISI = 60 ± 20 s and the dashed line represents the ratio for ISI = 40 ± 10 s that was used in the Budapest study (May et al., 2004). For high labilities, both curves approach unity. That is there are just as many possible decision points available to avoid a ns-SCR as there are to capture one relative to the background rate. However, at lower labilities, both protocols lead to many more opportunities to avoid a ns-SCR than to capture one. Because the stimulus types were determined randomly on a 50/50 basis, a differential measure of the proportion of ns-SCR rates between stimulus types cannot yield significant deviations. This was confirmed by a Stouffer's Z calculation for all cases in the Monte Carlo.

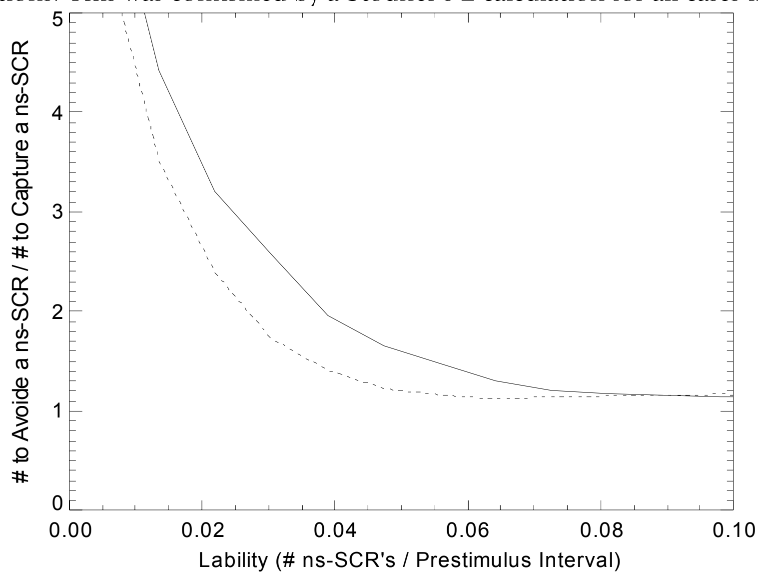


Fig. 1 Monte Carlo Opportunity Estimates

Since the Budapest protocol had a shorter ISI, we would expect and the Monte Carlo shows that the advantage favoring avoiding a ns-SCR would vanish more rapidly than for the longer ISI, because there is less time available between stimuli.

It is possible to infer from this calculation, that if DAT were operating, then it is more likely that the observed ns-SCR rate prior to controls might be less than that of the background. However, an equal contribution to a significant proportion difference might arise from both an increase prior to acoustic stimuli and a decrease prior to silent controls both relative to the background rate and still be consistent with a DAT interpretation.

Finally, if we assume that DAT-favorable decision possibilities are proportional to lability, then the DAT relationship above becomes:

$$Z = \sqrt{\frac{Lability_0}{Lability}}\, Z_0, \tag{1}$$

where $Lability_0$ is the lability leading to Z_0. Equation 1 may then be compared to the data.

Analysis and Results

In one prestimulus response study (May & Spottiswoode, 2003), 190 people were screened to obtain the required 100 for their formal study. The analysis here includes the data from all these people.

This section addresses the possible role of DAT circumstantially and from two different perspectives of supporting evidence.

Circumstantial Arguments against a Physiological Interpretation

It is well known that skin conductance responses to startle stimuli, such as loud noises or electric shocks, appear not to habituate as they do to more cognitive stimuli. If there were an anticipatory physiological response to a future startle stimulus then it too might not habituate.

The experimental result from the acoustic study gives the rate prior to acoustic stimuli as $P_1 = 0.091$ and prior to controls as $P_0 = 0.053$; therefore a response was a rather rare event. That is, on the average only one or two stimuli in each session might contain an anticipatory response. Figure 2 shows the effect size related to the proportion difference as a function of stimulus position.

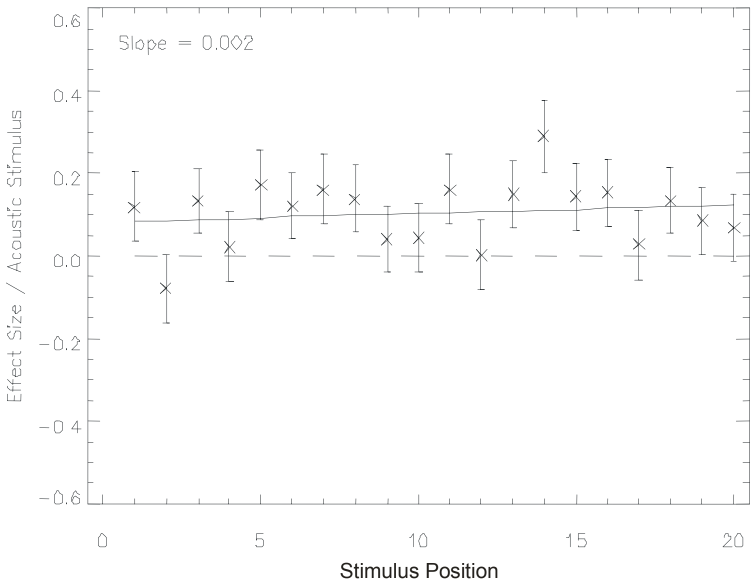


Fig. 2 Effect / Stimulus as a Function of Stimulus Position

The solid line is the weighted best fit to the data and the dashed line is the expectation value of zero for the effect size. The slope of the black line is 0.002 ES/(stimulus Position). On one hand, this appears to show no habituation over time as a physiological response might suggest, however, since the dependent variable is a proportion difference in a Bernoulli counting statistic and the events are rare, a slope near zero means that a rare response is just as likely to appear anywhere in the session. This does not support a usual physiological response mechanism.

A second circumstantial observation is that individuals who qualified to be included in the formal analysis were not known to have any psi ability in laboratory experiments with the exception of "star" consistent performers in the lengthy US Government-sponsored program. This later subgroup, however, did not perform according to their previous standards that were common in other kinds of laboratory studies.

Finally, even zombies (i.e., individuals who exhibit none or few ns-SCR's) respond, post stimulus, to startle stimuli and also show little or no habituation. Yet, these individuals show no anticipatory responses. This observation was so self-evident in pilot studies, that they were predefined as being excluded from the formal analysis.

The startle stimuli in these studies were limited with regard to their emotional impact; that is, compared to visual targets with high affectivity. Perhaps the underlying mechanism may be different for visual emotion targets.

Supporting Evidence for DAT

As described above, DAT suggests that it is likely to observe a ns-SCR rate prior to specific controls that is less then the background rate determined far from any overt stimuli. The ns-SCR rate prior to acoustic stimuli, P_1, prior to controls, P_0, and background rate, BK, was computed for each participant. Table 1 shows the results of paired t-tests for the white noise study (May & Spottiswoode, 2003) and for the study conducted in Budapest (May et al., 2004).

Table 1: DAT Analysis for Two Studies

	White Noise		Budapest	
Comparison	Paired-T(189)	P-Value	Paired-T(49)	P-Value
$P_1 > BK$	2.74	0.0032	0.163	0.436
$P_0 < BK$	-5.29	1.2×10^{-7}	-2.87	0.003

In both studies, the ns-SCR rate prior to formal controls was significantly below the general background rate. In the white noise study, the rate prior to future acoustic stimuli was significantly larger than the background rate; however, this is still consistent with a DAT interpretation.

Consider the 190 individuals in the white noise study. Each data record possessed a measured lability (i.e., the probability of a fully-formed ns-SCR in a 3.5-second interval far from stimuli). The labilities were combined in bins of 0.04 widths. The individual Z-scores within each bin were combined to form a Stouffer's Z. These Z's and associated 1-standard errors are shown as ×'s and error bars, respectively, in Figure 3 as a function of lability. A generalization of Equation 1, above, provides a fitting function given by:

$$y = a + \frac{b}{\sqrt{Lability}},$$

where a and b are free parameters for the fit. The solid line in Figure 3 represents the least-square's best fit to this functional form. The long-dashed lines represent the 1-standard errors for the fit. In general, given any data, it is possible to fit an arbitrary curve to those data. The fitting algorithm adjusts the parameters to provide the best fit possible.

The dashed thick line is a non-parametric representation of Equation 1–the DAT model. $Lability_0$ and Z_0 were the first paired points in the binned lability sequence so that the remaining curve is the prediction of Stouffer's Z's relative to the Z_0 and $Lability_0$ starting point. The data are noisy and, therefore, the χ^2 for the fit was 33.5 (df = 16, p = 0.006) which implies that the fit is not particularly good.[3] Although the lability dependence for the fit was set by DAT, the parameters *a* and *b* above were completely data-driven and could have been anything. Therefore even a less than ideal fit to the data remains strongly suggestive of DAT because of the near perfect agreement with the model as shown in Figure 3.

[3] Here χ^2 per degree of freedom was 2.07. The ideal fit is χ^2 per degree of freedom of one.

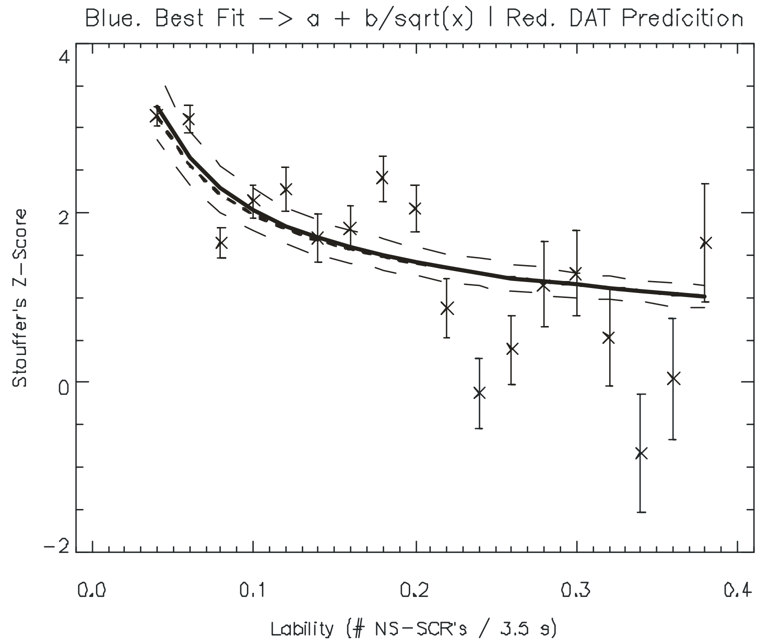


Fig. 1: DAT Analysis as a Function of Lability

We can compare this calculation with what might be expected under the hypothesis of an anticipatory physiological response. It is not unreasonable to expect that if participants were responding to future startle stimuli, they would do so regardless of their internal spontaneous ns-SCR rate. In other words, the Stouffer's Z would not be dependent upon lability and would appear as a horizontal straight line in Figure 3. A weighted straight line fit yielded a slope of -9.07 corresponding to 8.44 σ away from zero. Thus, a physiological model significantly fails to fit these data.

SUMMARY AND CONCLUSION

No retrospective analysis can be definitive; however, DAT predicts that with small ns-SCR rates, there will be an apparent suppression of the response rate prior to controls relative to the background rate, and in two separate studies that was confirmed significantly. For this finding to be an indication of a physiological response, the participants must first use their psi to know which 3.5 second interval will precede a silent control and then use some form of internal mechanism to prevent a ns-SCR. Perhaps both requirements might be possible to meet; however, the near-perfect fit of the Z-score dependency upon lability and that a linear assumption for a physiological fit to the data misses by more than 8 σ is additional, strong support for DAT as being the mechanism for this prestimulus response study and, perhaps, for all such studies.[4]

What is needed before a formal declaration can be made is a prospective, DAT-sensitive test.

ACKNOWLEDGEMENTS

We thank Richard Broughton, James Spottiswoode, and Zoltán Vassy for many helpful discussions. We also appreciate the financial support for the experimental studies by the Samueli Institute and Foundation BIAL.

[4] It is possible to speculate, based upon the DAT formalism above, that in studies where the participant is allowed to initiate each trial, that these results also may contain a significant DAT component.

References

Bierman, D. J., & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills, 84*, 689-690.

Hartwell, J. W. (1978). Contingent negative variation as an index of precognitive information. *European Journal of Parapsychology, 2*(2), 83-103.

Marsaglia, G., & Zaman, A. (1987). *Toward a universal random number generator* (No. FSU-SCRI-87-50): Florida State University.

May, E. C., Humphrey, B. S., & Hubbard, G. S. (1980). *Electronic systems perturbation techniques*: SRI International.

May, E. C., Paulinyi, T., & Vassy, Z. (2004). Audio prestimulus response: A replication, an extension and speculation on Decision Augmentation. *In preparation*.

May, E. C., Radin, D. I., Hubbard, G. S., Humphrey, B. S., & Utts, J. M. (1985). *Psi experiments with random number generators: An informational model.* Paper presented at the The Parapsychological Association 28th Annual Convention.

May, E. C., & Spottiswoode, S. J. P. (2003). *Skin conductance response to future audio stratle stimuli.* Paper presented at the The Parapsychological Association 46th Annual Convention, Vancouver, Canada.

May, E. C., Utts, J. M., & Spottiswoode, S. J. P. (1995). Decision augmentation theory: Toward a model for anomalous mental phenomena. *Journal of Parapsychology, 59*, 195-220.

Radin, D. I. (1997). *The Conscious Universe*. New York: Harper Collins.

Spottiswoode, S. J. P., & May, E. C. (2003). Skin conductance prestimulus response: analyses, artifacts and a pilot study. *Journal of Scientific Exploration, 17*(4), 617-641.

Stanford, R. G., & al., e. (1976). A study of motivational arousal and self-concept in psi-mediated instrumental response. *Journal of the American Society for Psychical Research* 70(2), 167-178.

Stanford, R. G., & Stio, A. (1976). A study of associated mediation in psi-mediated instrumental response. *Journal of the American Society for Psychical Research* 70(1), 55-64.

Stanford, R. G., Zenhausern, T. A., & Dwyer, M. A. (1975). Psychokinesis as psi-mediated instrumental response. *Journal of the American Society for Psychical Research* 69(2), 127-133.

Utts, J. M., & Heckard, R. F. (2004). Testing the difference between two population proportions. In *Mind on Statistics* (2nd ed., pp. 459-464). Belmont, CA: Brooks/Cole -Thomson Learning.

Vassy, Z. (1978). Method of measuring the probability of 1-bit extransensory information transfer between living organisms. *Journal of Parapsychology, 43*(2), 158-160.

Address for correspondence: Edwin C. May,
Laboratories for Fundamental Research,
415 Cambridge Ave., Suite 3, Palo Alto,
California 94306, USA.
E-mail: may@lfr.org

WHAT IS TO BE DONE: EVALUATING THE RITUAL HEALING THEORY[1]

James McClenon
Elizabeth City State University

ABSTRACT

The field of parapsychology has been criticized for not providing sufficiently replicable experiments. Researchers have found that their claims of achieving replicable experiments have not had the impact on mainstream science that might be expected. Content analysis of introductory psychology textbooks indicates that text authors tend to emphasize failure to replicate while ignoring evidence supporting parapsychological claims. The field of parapsychology demonstrates structural characteristics suggesting that it will not gain full legitimacy should it continue to follow present strategies.

The *Ritual Healing Theory* offers a means to bypass this problem. It is derived from research in the social psychology of anomalous experience, human genetics, and evolutionary psychology. It hypothesizes that dissociative abilities and the capacity for anomalous perceptions have physiological basis. It argues that early hominids developed dissociative capacities for coping with trauma. Later hominids devised therapeutic rituals that provided greater benefits to those with dissociative ability. Shamans conducted rituals that allowed those with dissociative abilities survival advantages and these individuals passed on dissociation genotypes more frequently. These processes caused the frequency of genes related to dissociation to increase. Experiences linked to dissociation (apparitions, waking and sleeping ESP, out-of-body experiences, and psychokinesis) were shaped and became more prevalent. These experiences generated beliefs in spirits, souls, life after death, and magical abilities, providing the foundation for shamanism, humankind's first religion. Shamanic healing practices, occurring over the last 30,000 years, continued selecting for dissociative and hypnotic genotypes, further shaping the physiological foundations for anomalous experience, religion, and ritual healing.

Ritual healing theory hypotheses are testable within the domains of psychical research, anthropology, archaeology, social psychology of religion, folklore, history, physiology, and medicine. Hypotheses include: (1) The extent of childhood trauma should be positively correlated with dissociative capacities. (2) Dissociative capacities should be positively correlated with propensity for anomalous experience. (3) Certain types of anomalous experience (apparitions, paranormal dreams, waking extrasensory perceptions, psychokinesis, and out-of-body experience) have biological basis and occur in all societies. (4) Certain types of experience, such as waking extrasensory perception and paranormal dreams have structural features associated with state of consciousness; waking extrasensory perceptions more often pertain to present events, lead to conviction, and provide less information than paranormal dreams which more often pertain to future events, fail to generate conviction, and provide greater information. These features are seemingly universal, implying physiological basis. (5) Propensity for anomalous experience is correlated with belief in spirits, souls, life after death, and magical abilities and certain experiences induce belief. (6) Researchers should find significant correlations between frequency of anomalous experience, temporal lobe signs, transliminality, and cognitive openness, traits hypothesized to be part of the shamanic complex. (7) Capacity for ritual healing should be positively correlated with dissociative/hypnotic ability. (8) Anthropologists can observe ritual healing processes through field studies. They can detect patterns indicating a shamanic complex. (9) Historians of medicine can document that all ancient medical practices contain rituals based on placebo and hypnotic processes.

INTRODUCTION

The ritual healing theory offers an alternative to parapsychological paradigms. Parapsychologists struggle with the demand to provide a fully replicable experiment. Skeptics have erected rhetorical obstacles; they

[1] This paper is a revision of a forthcoming chapter – The Ritual Healing Theory: Hypotheses for Psychical Research, In M. A. Thalbourne and L. Storm (Eds.). *Parapsychology in the 21st Century: The Future of Psychical Research*, Jefferson, NC: McFarland. It is submitted with the permission of the editors and future publisher.

argue that exceptional claims require exceptional proof and that parapsychological claims are so exceptional that present "proofs" are insufficient. Patterns within the history of science suggest that parapsychologists will find it difficult to overcome these arguments. In the past, successful innovative scientific fields have grown rapidly during their decade of origin (Griffith & Mullins, 1972; McClenon, 1984). This allowed these fields to use their political power to overcome rhetorical obstacles. The Parapsychological Association, the field's major scientific organization, has not grown markedly. Membership was 205 in 1970, 279 in 1980, 306 in 1983, 275 in 1986, 246 in 1992, 251 in 1999, and 254 in 2001 (McClenon, 1994: 200; Varvoglis, 1999; Utts, 2001). The history of science indicates that successful scientific paradigms resolve central research questions, overcoming the objections of their critics. Parapsychology has not been able to do this. Analysis of the treatment of parapsychology within introductory psychology texts indicates continuing skepticism, lasting over many decades (Roig, Icochea, & Cuzzucoli, 1991; McClenon, et al, 2003). Parapsychology's longevity, non-scientific financial support, and stagnant growth illustrate a "deviant science" (McClenon, 1984), suggesting that future claims of experimental replicability will also be treated skeptically.

This paper describes "what is to be done" to overcome obstacles to parapsychological legitimacy. As with political revolts, scientific revolutions require ideologies around which collective movements can be organized. The ritual healing theory provides an ideology allowing this process. It does not involve "extraordinary claims" but encourages research regarding anomalous perceptions within accepted evolutionary psychology paradigms. It provides a wide range of testable hypotheses. Findings within the suggested paradigm can be published in journals dealing with religion, mind-body medicine, evolutionary psychology, and anomalous experience.

Evolutionary psychology has gained increasing support among anthropologists and psychologists during the past few decades. The field generates hypotheses pertaining to a wide variety of topics ranging from mating strategies to altruism. Although various introductory psychology textbooks focus specifically on evolutionary psychology (Buss, 1998; Gaulin and McBurney, 2001; Palmer and Palmer, 2002), all major introductory psychology texts review evolutionary arguments regarding genetic influence of human behavior.

Behavioral geneticists argue that all behavior involves gene/environment interplay (McGuffin, Riley, & Plomin, 2001). Researchers are often surprised at the important role that genes play. For example, studies indicate that about 50% of the observed variance of measures of religious interest, attitudes, and values are genetically influenced (Waller et al., 1990).

The most solid genetic findings about individual differences in human behavior come from quantitative genetic research such as twin and adoption studies that consistently converge on the conclusion that genetic variation makes a substantial contribution to phenotypic variation for all behavioral domains. The best-studied areas are psychopathology, personality, and cognitive abilities and disabilities. There are two striking findings. The first is that nearly all behaviors that have been studied show moderate to high heritability--usually, to a somewhat greater degree than do many common physical diseases. Second, although environment plays a role, its contribution tends to be of the nonshared type, that is, environmental factors make people different from, rather than similar to, their relatives (McGuffin, Riley, & Plomin, 2001: 1232).

Ultimately, geneticists will identify impacts (or lack of impact) of the 30,000 human genes. Any trait that is universal consitutes a phenotype, a characteristic with genetic basis. For example, if the capacity for dissociation is universal, then this trait will be found to have genetic basis. Social scientists can contribute to this research through identifying universal features.

The Darwinian paradigm allows a theory explaining the origin of religious sentiment. The ritual healing theory argues that ancient hominids with dissociative capacities had survival advantages since they were better able to cope with the psychological effects of trauma. As dissociative genotypes increased, hominids devised therapeutic rituals using these capacities, further increasing the frequency of dissociative genotypes (McClenon, 1997a, 2002a). Eventually, Homo sapiens linked rituals with language, coupling suggestions with altered states of consciousness (ASC). As a result, trance rituals shaped human hypnotizability, a genetically-based trait (Duke, 1969; Morgan, 1973; Morgan, Hilgard, & Davert, 1970). This process also shaped spontaneously occurring anomalous experiences, linked to dissociation. Such experiences, including apparitions, waking and sleeping extrasensory perceptions (ESP), psychokinesis (PK), sleep paralysis, out-of-body experiences (OBE), and near-death experiences (NDE), generated beliefs in spirits, souls, life after death, and magical abilities. These beliefs provided the foundation for shamanism. Shamanism was humankind's first religion, practiced by all hunter-gatherer societies (Winkelman, 1992, 2000). As a result, folk religious traditions regarding spirits, souls, life after death, and magical abilities exist in all large societies (McClenon, 1997a, 2002a).

Shamanism is defined as a religious system in which practitioners go into trance to contact spirits thought to affect living people. Shamanism can be operationally defined as including specific characteristics identified in the anthropological literature (Winkleman's (1992: 48). Paleolithic rock and cave art, indicating use of altered states of consciousness, suggests that shamanism has been practiced for over 30,000 years (Lewis-Williams & Dowson, 1988). The existence of ritual and hypnotic processes in animals implies that hominids gained an increasing capacity to use altered states of consciousness in ritual during human evolution (Hoskovec & Svorad, 1969; McClenon, 2002).

Evolutionists note that physiological change can occur within relatively brief time spans. Lumsden and Wilson (1983: 152) argue that the coupling of genes and culture drove human evolution in a rapid manner. They argue that significant genetic changes have occurred within a mere 50 generations, or approximately 1,000 years. Human teeth and bone structure changed with advances in Paleolithic technology, illustrating the effect of culture on human evolution. Increasing lactose tolerance, associated with human domestication of cattle, occurred within the past 10,000 years.

The ritual healing theory does not argue that humankind gained direct benefits from psi. It does not pass judgement on the authenticity of psi or other exceptional claims. It argues that certain anomalous experiences have universal features reflecting their physiological bases and that belief in ideologies associated with these experiences facilitate therapeutic rituals. These rituals provide greater benefits to those with dissocative/hypnotic capacities.

Dawkins (1999) explains the evolutionary orientation by noting that genes are like oarsmen in a racing boat. After a number of races, the winning oarsmen (genes) are randomly assigned to new boats (bodies) and the races continue. All boats contain incompetent oarsmen but there is a tendency for skillful teams to win. Genotypes, groups of genes with a specific purpose, are like oarsmen teams fulfilling complex functions. Such teams are often effective only in certain environments. The ritual healing theory argues that oarsmen allowing dissociation and hypnosis proved valuable during humankind's era of evolutionary adaptation. These winning dissociation/hypnosis teams contained members facilitating trance, apparitions, ESP, PK, OBE, NDE, and other anomalous perceptions sustaining shamanism. Ultimately, oarsmen teams were shaped to allow the modern propensity for religion.

Irwin (1992, 1993) provides a model hypothesizing relationships between psychological variables and cultural factors of particular interest to parapsychologists. His model can be modified to coincide with the evolutionary paradigm and to portray the ritual healing theory (see Figure 1). Within this model, "encouragement of childhood fantasy" and "childhood trauma" affect "dissociative ability." Paleolithic people (particularly children) with dissociative ability had greater survival advantages when exposed to trauma than those lacking this ability. Over time, genotypes associated with dissociation increased. These genotypes are linked to various anomalous experiences that generate beliefs providing the ideological foundation for shamanism (McClenon, 1997a, 2002a). Because shamanism involved therapeutic rituals that provided greater benefits to those with dissociative capacities, the frequency of dissociative genotypes continued to increase and be shaped by ritual process.

Using Dawkin's (1999) metaphor, dissociation genotypes contain oarsmen allowing anomalous experiences. Over time, the compositions of these teams were shaped by evolutionary selection, governed by ritual suggestion. Genotypes (oarsmen teams within each boat) related to dissociation/hypnotizability increased because they contributed to survival (the boat's victory). Because oarsmen on these teams facilitated trance and other anomalous experiences, shamanism developed in all hunter-gatherer societies.

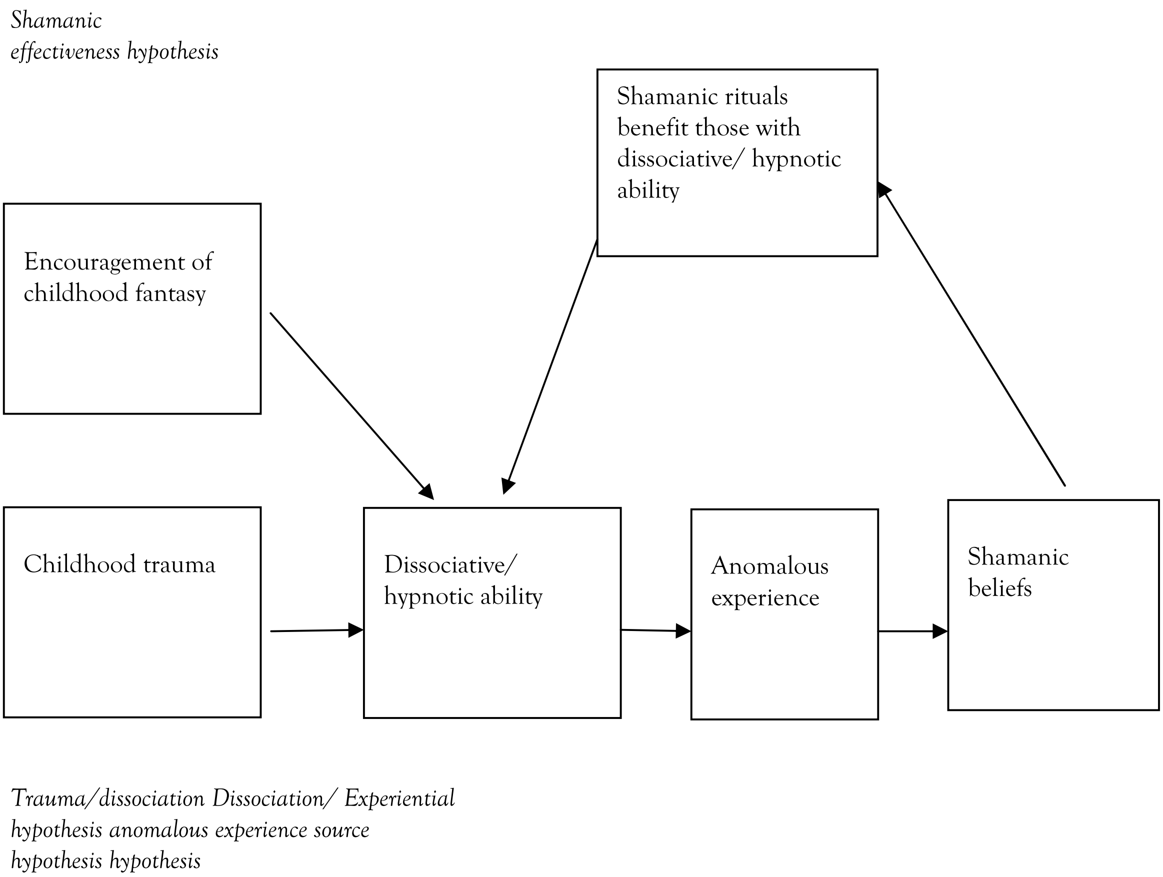


Figure 1: Ritual Healing Theory

Dissociative/hypnotic capacities could not increase infinitely because these traits had negative consequences. Highly dissociative people are fantasy prone and tend to suffer from psychosomatic and dissociative identity disorders. Not all such people recover when treated by shamanic ritual. The ritual healing theory portrays a process where those with moderate levels of dissociative/hypnotic capacity have optimal survival advantage.

The ritual healing theory can be evaluated within the context of four broad hypotheses: (1) The trauma/dissociation hypothesis specifies a positive correlation between childhood trauma and dissociative/hypnotic capacities (2) The dissociation/anomalous experience hypothesis specifies positive correlations between dissociative/hypnotic processes and incidence of anomalous experiences. (3) The experiential source hypothesis specifies that certain anomalous experiences instill belief in spirits, souls, life after death, and magical abilities. (4) The shamanic effectiveness hypothesis argues that ritual healing can be effective due to hypnotic and placebo effects.

THE TRAUMA/DISSOCIATION HYPOTHESIS

This hypothesis specifies that incidence of childhood trauma and neglect are correlated with capacity for dissociation and hypnosis. It is almost axiomatic within the therapy literature that traumatic events cause some children to develop dissociative capacities (Albini and Pease, 1989; Fink, 1988; Putnam, 1985; Spiegel, 1986). Although excessive use of dissociation results in identity disorders (Ross, 1997), dissociation is not intrinsically pathological but is related to fantasy and imaginative ability (Putnam, 1991). The "dissociation benefit" argument, used within the ritual healing theory, is based on two assumptions: (1) Dissociation can protect against negative effects of stress. (2) Humans used this mechanism for sufficient time that related genes become more widespread.

Terms within the above hypotheses have been defined and operationalized. Dissociation is defined as "experiences and behaviors that exist apart from, or have been disconnected from, the mainstream of one's conscious awareness, behavioral repertoire, and/or self-concept" (Krippner, 1994: 357). Hypnosis is "a psychophysiological condition in which attention is so focused that there occurs a relative reduction of both peripheral awareness and critical analytic mentation, leading to distortions in perception, mood, and memory which in turn produce significant behavioral and biological changes" (Wickramasekera, 1987: 12). Standardized scales allow evaluation of childhood trauma (Kent & Waller, 1998; Sanders & Becker-Lausen, 1995), capacity for dissociation (Lynn and Rhue, 1994) and hypnotic suggestibility (Fromm and Nash, 1992).

The ritual healing model recognizes that dissociative/hypnotic genotypes have been shaped by both natural and cultural factors (such as "encouragement of childhood fantasy") and that, as a result, "nature" and "nurture" cannot be completely differentiated. Modern evolutionary psychologists recognize that ecology and culture are linked since both elements shaped evolutionary selection.

Frequency of anomalous experience should be positively correlated with childhood trauma and neglect -- with dissociative capacity as a mediating variable. The ritual healing theory does not specify a causal linkage. Dissociation does not necessarily *cause* anomalous experience and the reason for the hypothesized correlation is not specified. Dissociative genotypes may include genes facilitating anomalous experience and similar cognitive processes may generate dissociative and anomalous perceptions.

The term *anomalous experience* refers to specific forms of perception, having similar identifying features in all societies. Anomalous experiences include apparitions, waking and sleeping ESP, PK, OBE, NDE, and sleep paralysis. Scales measuring the incidence of these perceptions are described in peer- reviewed journal articles (McClenon, 1994, 2002; Kumar and Pekala, 2001).

Psychical researchers can evaluate hypotheses specific to their areas of expertise: (1) Groups claiming higher levels of anomalous experience should report higher rates of childhood trauma and neglect. (2) Samples of psychic practitioners and spiritual healers, who tend to use trance and dissociative techniques, should report higher levels of childhood trauma and neglect than general populations.

Ethologists (observers of animal behavior) might investigate additional hypotheses. The ritual healing theory predicts that dissociative and hypnotic capacities exist in alternate forms in various animals. Dissociative capacities aid these animals when confronted with certain stressful stimuli. For example, some animals react to threats using the Totstell reflect, often called "animal hypnosis" since it is parallel to human hypnosis (Hoskovec & Svorad, 1969). These animals "play dead," adapting unusual forms of consciousness in order to avoid predators' attacks. Such startle responses include rapid changes in consciousness, paralysis, and the "sleepy" appearance associated with human hypnosis. Ethologists note that repetitive, non-verbal rituals function as "hypnotic" inductions for some animals. Völgyesi (1966) induced trance in primates using non-verbal, repetitive techniques suitable for humans.

THE DISSOCIATION/ANOMALOUS EXPERIENCE HYPOTHESIS

The ritual healing theory argues that: (1) Researchers should find significant correlations between frequency of anomalous experience, dissociative experience, temporal lobe signs, transliminality, and cognitive openness (variables that might be regarded as part of a "shamanic complex"). (2) Researchers can observe universal features within spiritual healing rituals related to these variables.

Much evidence supports these hypotheses. Research indicates that the propensity for anomalous experience, hypnosis, dissociation, fantasy proneness, temporal lobe lability, and thinness of cognitive boundaries are inter-correlated (McClenon, 1994, 2002a; Targ, Schlitz, & Irwin, 2000). This literature supports the argument that a “shamanic complex” governs the nature and incidence of shamanism.

(1) Dissociation and Paranormal Experience

Various researchers have found positive significant correlations between dissociative capacity and frequency of paranormal and anomalous experience (Pekala et al., 1995; Ross & Joshi, 1992; Ross et al. 1991; Richards, 1991). Richards’ (1991) findings are typical. He concluded that the dissociation was most correlated with waking clairvoyance, precognition, apparitions, psychokinesis, and volitional telepathy. He argues that out-of-body experiences, trance channeling, and contact with spirit guides imply dissociative processes (as indicated by high significant correlations).

(2) Hypnotizability and Paranormal Experience

Kumar and Pekala (2001: 275-276) summarize findings regarding hypnotizability and paranormal experience: “A total of 23 correlations were reported in 11 different studies.... three were reported as nonsignificant (-.20, p> .05; .13, p>.001; [and one did not report an r value but reported that it was non-significant]; the rest varied between .17 and .55 (all at least p<.05). The median correlation (excluding the negative value) was .31. Studies.... examining group differences in experiences also support a relationship [except for one study] in the sense that participants with high hypnotizability tend to report a greater number of experiences than those with low hypnotizability.”

(3) Shamanic Complex

Other variables, correlated with each other, hypnosis, and dissociation, have been found significantly correlated with frequency of anomalous experience. Thalbourne and Delin (1994, 1999) conducted a principal-components analysis of variables such as belief in paranormal phenomena, magical ideation, manic and depressive experiences, and scores on a creative personality scale. They found that a single factor accounted for 52.5% of the variance in one study and 54.2% in a replication. They labelled this factor “transliminality,” the degree to which there is a gap in the barrier or gating mechanism between the unconscious (subliminal) and conscious mind (Thalbourne, et al., 1997). Transliminality was highly correlated with a measure of mysticism, and people who were high in transliminality were more susceptible to incursions of idealitional and affective input from subliminal regions (Thalbourne & Delvin, 1999).

Hartmann (1991) conducted parallel studies, linking cognitive processes with anomalous experience. He hypothesized positive correlations between a measure of “thinness” of cognitive boundaries and factors equivalent to transliminality. He defined cognitive boundaries as barriers to the spontaneous flow of images and information within the brain. People with thin boundaries have the sense of merging with their perceptions. They reveal greater fluidity of thoughts and feelings since they have fewer barriers or walls separating them cognitively from the world. Thin cognitive boundaries allow hypnotic suggestions to affect unconscious processes, a characteristic associated with certain pathologies. Thinness facilitates the flow of anomalous perceptions into conscious awareness; as a result, those revealing thin boundaries on Hartmann’s scale tended to report more frequent anomalous experiences and to be more hypnotizable (Hartmann, 1991).

Persinger and his associates conducted a series of studies indicating that responses to questionnaire items related to temporal lobe epilepsy (temporal lobe signs) are related to specific EEG patterns indicating temporal lobe lability (Kakarec & Persinger, 1990). These responses correlated with the propensity to report anomalous and religious experiences (Persinger, 1984a, 1984b, Persinger & Makarec, 1987, 1993; Persinger & Valliant, 1985). This body of evidence supports the argument that the propensity for anomalous experiences has physiological bases and is linked to other cognitive processes.

Studies linking frequency of anomalous experience with other parameters support arguments for the existence of a shamanic complex (McClenon, 1994, 2002a). Anthropologists observe that people with a propensity for anomalous experience often suffer from psychologically based disorders and culturally specific pathologies (often attributed to spiritual forces). Such people may be healed by spiritual practitioners and, as part of this process, become spiritual practitioners themselves (Lewis, 1971). Shamanic healing rituals include elements involving hypnotic induction, suggestion, and changes in self-concept that seem designed to benefit those scoring high on shamanic complex variables.

EXPERIENTIAL SOURCE HYPOTHESIS

The ritual healing theory hypothesizes that physiologically-based anomalous experiences generate beliefs in spirits, souls, life after death, and magical abilities. Experiential source hypotheses are particularly amenable to evaluation by psychical researchers. Psychical researchers could replicate previous findings regarding anomalous experience: (1) Collections of anomalous accounts from any society should include stories regarding apparitions, waking ESP, paranormal dreams, psychokinesis, out-of-body experience, sleep paralysis, synchronicity, and spiritual healing. Universal features within these accounts imply physiological bases. (2) Apparitions, waking ESP, and paranormal dreams have structural features related to recurring elements and altered states of consciousness associated with family, death, temporal factors, and conviction, a finding implying physiological basis for these episodes (3) People who report one experience are more likely to report multiple experiences and to reveal particular beliefs and behaviors as a result. Although culture shapes experiencers' interpretations and reporting of any particular experience, experiential source hypotheses specify that physiologically-based, universal elements within accounts shape folk religious beliefs in a common direction in all societies.

Hufford (1982) developed early experiential source arguments. He described a dominant social scientific assumption that he termed the "cultural source theory." This orientation argues that anomalous accounts are merely cultural products, having no basis beyond the experiencer's own belief system. He set forth an alternate paradigm, the experiential source theory, arguing that some forms of anomalous experiences generate and shape folk religious traditions, independent of culture. Hufford's (1982) findings regarding sleep paralysis strongly supported the experiential source position.

McClenon's (1994) analyses of apparitions, waking ESP, paranormal dreams, out-of-body experiences, psychokinesis, synchronicity, and other anomalous episodes extended Hufford's arguments. These forms of anomalous experience reveal similar defining elements in all societies, bringing about parallel beliefs in spirits, souls, life after death, and magical abilities (McClenon, 1994, 2002a). For example, people who link an apparition or psychokinesis to a deceased person often come to believe in an after-life. Those who report an out-of-body experience are more likely to accept the existence of souls. Near-death experiences are particularly powerful for inducing belief in an after-life among non-believers. Those who report ESP episodes are more likely to believe in magical abilities. Although surveys establish correlations between experiences and belief, the direction of causality can be better investigated through interviews and participant observation.

Analyses of survey responses from Japan, China, Europe, and the USA reveal that all groups report ESP and contacts with the dead (McClenon, 1994). Collections of narrative accounts from Finland, Germany, Great Britain, and the USA indicate that people from different cultures report similar forms of apparitions, waking ESP, paranormal dreams, psychokinesis, out-of-body experience, and synchronicity (McClenon, 1994). Anomalous stories were classified into clearly defined categories using reliable coding systems (McClenon, 1994, 1997b, 2000). The data implied that these forms of anomalous experience have physiological basis, just as do dreams and trance perceptions. Dominant social scientific positions arguing that anomalous experiences are totally produced by culture are clearly false.

Studies also indicated that waking extrasensory perceptions, paranormal dreams, and apparitions have inherent structural features, consistent among cultures (McClenon, 2000). This evidence implies physiological bases for these cognitive events. Analysis of collections from Finland, Germany, China, and the USA revealed that: (1) ESP tends to pertain to family members. (2) Death is often an important theme within ESP accounts. (3) Paranormal dreams tend to pertain to future events while waking ESP tends to pertain to present events. (4) Waking ESP tends to generate greater conviction, indicated by the respondent taking action, than do paranormal dreams. (5) Paranormal dreams tend to provide more information than do waking ESP episodes. (6) There is a tendency for "quality of information" to be negatively correlated with "severity of event" within paranormal dreams. Paranormal dreams often fail to reveal the identity of a person who died, while providing more complete information for events not associated with death. (7) Apparitions contain similar "abnormal features of perception" in all societies. These features include disappearance of image, insubstantial image, glowing image, white or black clothes, sickly or deformed image, partial body, abnormal walking or floating, and abnormal sound (Emmons, 1982; McClenon, 1994).

Correlational studies verify links between anomalous experience and belief (Pekala, Kumar, and Cummings, 1992; Targ, Schlitz, and Irwin, 2000) and qualitative studies indicate that many people report that anomalous experiences create specific beliefs (McClenon, 1994). Although a dominant paradigm within religious studies portrays anomalous experiences as products of belief, tests of this argument consistently demonstrate it to be false (Hufford, 1982; McClenon, 1994). For example, historical analysis of Icelandic mediumship revealed that psychic experiences and performances transformed Icelandic religious traditions in an innovative direction (Swatos and Gissurarson, 1997).

Social-psychological research indicates that attitudes formed by direct experience are stronger than those gained through other means, and that experience-based attitudes are better predictors of later behavior than attitudes gained through other means (Millar and Millar, 1996). People reporting frequent anomalous experiences reveal particularly robust belief-systems that we would expect to affect their behavior. Researchers have found that anomalous experiences are not distributed normally – that the majority of experiences are reported by a small segment of any population. Such people tend to hold powerful convictions regarding spirits, souls, life after death, and magical abilities (Greeley, 1995, 1987; McClenon, 1994; Palmer, 1979).

Psychical researchers can engage in participant observation, designed to monitor creation of belief through experience. Some methods allow semi-controlled conditions. "Sitter groups" investigate PK through table tipping experiments (Batcheldor, 1966, 1979, 1984; Owen & Sparrow, 1976). These groups sit regularly, often once a week, for an hour or more, with their hands on a table, seeking to generate psychokinetic effects. Groups often report anomalous table movements and unexplained auditory "rapping" thought to be generated by spirits. Although observational results have bearing on parapsychological theories, it is often impossible to verify psi within these settings. Whether authentic or not, people report that "sitter group" psi affects their beliefs.

The processes by which sitter groups generate belief through experience are similar to processes observed by anthropologists within shamanic groups. Shamanic, spiritualist, and sitter groups seem to experience similar forms of psychic phenomena even though cultures vary (McClenon, in press). Participants report seeing spirit lights, perceiving objects moving magically, hearing unexplained sounds and voices, and even feeling the whole room shake as during an earthquake. Psychical researchers could contribute to the anthropological literature regarding shamanism, anomalous experience, and belief (Young and Goulet, 1994). Such evidence refutes cultural source hypotheses and supports experiential source hypotheses.

Ethnographers can conduct haunting and poltergeist investigations within this paradigm. They should find that haunting and poltergeist experiences occur more often among those reporting previous anomalous experiences, that haunting accounts contain culturally universal features, and that resulting stories contribute to recurring features with folk religious traditions (McClenon, 2001).

THE SHAMANIC EFFECTIVENESS HYPOTHESIS

The ritual healing theory argues that shamanic ritual provides benefits to clients through hypnotic and placebo effects. It does not specify that paranormal events are impossible or that hypnotic and placebo effects are the *only* benefits that ritual healing provides. The shamanic effectiveness hypothesis specifies that people with shamanic complex traits tend to gain greater benefits from shamanic rituals.

Much evidence supports the shamanic effectiveness hypothesis. It is almost axiomatic within anthropology that spiritual healing can be effective due to psychological processes (Bergman, 1973; de Montellano, 1975; Finkler, 1985; Garrison, 1977; Harner, 1973; Kapferer, 1983; Kleinman and Sung, 1979; Kleinman, 1980; Laderman, 1987, 1991; Lambo, 1974; Moerman, 1979; Sharon, 1978; Vogel, 1970). This argument coincides with findings from the emerging fields of psychoneuroimmology and mind-body medicine: health is influenced by psychological states affected by suggestion (Benson, 1996; Friedman, Klein, & Friedman, 1996).

Much evidence implies that spiritual healing effectiveness is due, in part, to hypnotic and placebo effects (McClenon, 1997a,b, 2002a). Ritual healing symptoms, procedures, and outcomes parallel those associated with hypnotic processes. Rituals induce altered states of consciousness (ASC) through sensory restriction or overload, fasting, ingesting drugs, repetitive movements, dancing, drumming, chanting, prayer, and prolonged postures. ASC, coupled with suggestion, typically constitute hypnosis. Cognitive states associated with shamanism are linked to hypnosis: alterations in thinking, changes in sense of time and body image, loss of control, changes in emotional expression, perceptual distortions, changes in meaning and significance, sense of ineffability, feelings of rejuvenation, and hypersuggestivity (Ludwig, 1966).

Paleo-anthropological evidence implies a link between ritual and ASC. This should be expected since repetitive rituals tend to induce ASC. Specific features within Paleolithic cave paintings, produced over 30,000 years ago, indicate that ancient peoples induced ASC (Lewis-Williams and Dowson, 1988). Paleolithic people, over 50,000 years ago, cared for their sick and engaged in symbolic actions seemingly for their benefit. Shamanism, having physiological basis, provided the foundation for all later religious forms (Winkelman, 1992, 2000). The ritual healing scenario encompasses sufficient time that shamanism would have affected the frequency of genotypes related to this process. With the invention of writing, humans left evidence connecting ritual healing and hypnosis. Ancient texts provide "abundant evidence which shows that hypnosis or a similar induced ASC was used in ancient Greece, Egypt, India, China, Africa, and pre-Columbian America" (MacHovec, 1975:215).

A content analysis of anomalous experience narratives implies that, among the various forms of experience, spiritual healing has had the greatest impact on human evolution. Judges evaluated 1215 anomalous experience narratives, determining the degree each account indicated positive emotion (which implies evolutionary benefit) and "direct benefit." Among the experience types, spiritual healing accounts were most associated with evolutionary benefits while waking and sleeping ESP were often linked to social costs (McClenon, 2002b).

Various studies indicate relationships between dissociation, ritual, and healing. Krippner (1994) reviewed studies of patients with dissociative identity disorders who benefited from ritual treatments and portrayed dissociative processes within spiritual healing. Krippner and Colodqin (1989) noted the use of Native American and Oriental healing methods to treat combat veterans with posttraumatic stress disorders. Goodwin, Hill, and Attias (1990) encouraged psychotherapists to familiarize themselves with historical and folk techniques of exorcism since these techniques might be adapted to treat dissociative disorders. nthropologists describe processes within spiritual healing that imply hypnosis and dissociation (McClenon, 2002a). Csordas (1997), for example, linked successful spiritual healing among charismatic Christians with transformations of identity. The processes he described seem related to the shamanic complex. A person suffering from psychosomatic symptoms may be healed through gaining a healthy identity. Psychical researchers can contribute to this literature by documenting the types of individual that are more likely to benefit from magical therapeutic performances.

Symptoms treated through spiritual healing and hypnosis coincide. Clinical studies indicate that hypnosis is particularly effective for pain, asthma, warts, headache, burns, bleeding, gastrointestinal disorders, skin disorders, insomnia, allergies, psychosomatic disorders, and minor psychological problems (Bowers and LeBaron, 1983; Brown, 1992). Folk healing methods also effectively deal with these problems, often through suggestion. Researchers note that hypnotic suggestion does not require trance induction to be effective. Hypnosis can change the response of human skin to heat, probably through reducing edema and fluid retention following thermal injury (Margolis et al. 1983). It can also accelerate healing – perhaps through mechanisms involving hypnotic control of blood flow (Chapman, Goodell, and Wolff, 1959; Ullman, 1947, Moore and Kaplan, 1983: Barber, 1984). This process may explain some of the extremely anomalous healing stories found in all societies (McClenon, 2002a). For example, patients may cut off blood flow to cancerous tumors as a result of hypnotic suggestion, causing cancerous growths to wither away.

The argument that spiritual healing generates hypnotic and placebo effects does not preclude the possibility that unexplained processes may also occur. The ritual healing theory takes no position regarding the existence of magical or psi effects. Psychic researchers have investigated various unusual phenomena associated with shamanism and spiritual healing. Such phenomena include alleged extrasensory perception, psychokinesis, and firewalking. The ritual healing theory argues that such perceptions contribute to hypnotic and placebo healing since they inspire belief. Psychical researchers can document the effects of these performances on people's belief. People witnessing and accepting psychic performances are predicted to benefit more from associated therapeutic suggestions than those not exposed or believing. Healed people should reveal more dissociativity, hypnotizability, frequency of anomalous experience, temporal lobe lability, and thinner cognitive boundaries than those not healed.

CONCLUSIONS

The ritual healing theory provides a variety of hypothesis that psychical researchers can test. These hypotheses pertain to spontaneous anomalous experience, spiritual healing, and anomalous performance. The ritual healing theory allows researchers to transcend conflicts between skeptics and believers regarding the authenticity of psi, and facilitates the investigation of anomalous claims within mainstream scientific paradigms. This orientation can contribute to the scientific study of religion and lead to more effective treatments of psychologically influenced disorders.

References

Albini, T. K. & Pease, T. E. (1989). Normal and Pathological Dissociations of Early Childhood. *Dissociation, 2,* 144-150.

Barber, J. (1984). Changing 'Unchangeable' Bodily Processes by (Hypnotic) Suggestions: A New Look at Hypnosis, Cognitions, Imagining and the Mind-Body Problem, in A. A. Sheikh, A.A. (Eds.) *Imagery and healing* (pp. 69-128). Farmingdale, NY: Baywood Publishing Co.

Batcheldor, K. J. (1966). Report on the Case of Table Levitation and Associated Phenomena. *Journal of the Society for Psychical Research, 43,* 339-356.

Batcheldor, K. J. (1979). PK in Sitter Groups. *Psychoenergetic Systems, 3,* 77-93.

Batcheldor, K. J. (1984). Contributions to the Theory of PK Induction from Sitter-Group Work, *Journal of the American Society for Psychical Research, 78,* 105-132.

Benson, H., with Stark, M. (1996). *Timeless Healing: The Power and Biology of Belief,* New York: Scribner.

Bergman, R. L. (1973). A School for Medicine Men. *American Journal of Psychiatry, 130,* 663-66.

Bowers, K.S. & LeBaron, S. (1986). Hypnosis and Hypnotizability: Implications for Clinical Intervention. *Hospital and Community Psychiatry, 37,* 457-467.

Brown, D. P. (1992). Clinical Hypnosis Research Since 1986, In E. Fromm and M. R. Nash (Eds.), *Contemporary Hypnosis Research* (pp. 427-458), New York: Guilford Press.

Buss, D. M. (1998). *Evolutionary Psychology: The New Science of the Mind.* Boston: Allyn and Bacon.

Chapman, L., Goodell, H., & Wolff, H. (1959). Increased Inflammatory Reaction Induced by Central Nervous System Activity. *Transactions of the Association of American Physicians, 72,* 84-110.

Csordas, T. J. (1997). *The Sacred Self: A Cultural Phenomenology of Charismatic Healing.* Berkeley, CA: University of

California Press.

Dawkins, R. (1999). *The Selfish Gene*. New York: Oxford University Press.

de Montellano, B. O. (1975). Empirical Aztec Medicine. *Science, 188,* 215-20.

Duke, J. D. (1969). Relatedness and Waking Suggestibility. International Journal of Clinical Experimental Hypnosis, 17: 242-50.

Emmons, C. F. (1982). Chinese Ghosts and ESP: A Study of Paranormal Beliefs and Experiences. Metuchen, NJ: Scarecrow Press.

Fink, D. L. (1988). The Core Self: A Developmental Perspective on the Dissociative Disorders. *Dissociation, 1*, 43-47.

Finkler, K. (1985). *Spiritualist Healers in Mexico: Successes and Failures of Alternative Therapeutics*. New York: Bergin and Garvey Publishers.

Friedman, H., Klein, T. W., & Friedman, A. L. (Eds.). (1996). *Psychoneuroimmunology, Stress, and Infection*, Boca Raton: CRC Press.

Fromm, E. & Nash, M. R., Eds. (1992). *Contemporary Hypnosis Research*, New York: Guilford Press.

Garrison, V. (1977). The "Puerto Rican Syndrome" in Psychiatry and Expiritismo. In V. Crapanzano & V. Garrison (Eds.), *Case Studies in Spirit Possession* (pp. 383-449). New York: John Wiley and Sons.

Gaulin. S. J. C. & McBurney, D. H. (2001). *Psychology: An Evolutionary Approach*. Upper Saddle River, NJ: Prentice Hall.

Goodwin, J., Hill, S., & Attias, R. (1990). Historical and Folk Techniques of Exorcism: Applications to the Treatment of Dissociative Disorders. *Dissociation, 3,* 94-101.

Graffin, N. F., Ray, W. J., & and Lundy, R. (1995). EEG Concomitants of Hypnosis and Hypnotic Susceptibility. *Journal of Abnormal Psychology, 104*, 123-131.

Griffith, B. C. & Mullins, N. C. (1972). Coherent Social Groups in Scientific Change. *Science 177* (Sept. 19), 959-964.

Greeley, A. M. (1975). *Sociology of the Paranormal: A Reconnaissance*. Beverly Hills, CA: Sage Publications.

Greeley, A. M. (1987). Mysticism Goes Mainstream. *American Health, 6,* 47-49.

Harner, M. J. (1973). *Hallucinogens and Shamanism*. London: Oxford University Press.

Hartmann, E. (1991). *Boundaries in the Mind: A New Psychology of Personality*. New York: Basic Books.

Hoskovec, J. & Svorad, D. (1969). The Relationship Between Human and Animal Hypnosis, American Journal of Clinical Hypnosis, 11, 180-182.

Hufford, D. J. (1982). The Terror that Comes in the Night: An Experience-Centered Study of Supernatural Assault Traditions. Philadelphia: University of Pennsylvania Press.

Irwin, H. J. (1992). Origins and Functions of Paranormal Belief: The Role of Childhood Trauma and Interpersonal Control. *Journal of the American Society for Psychical Research, 86*, 199-208.

Irwin, H. J. (1993). Belief in the Paranormal: A Review of the Empirical Literature. *Journal of the American Society for Psychical Research, 87*, 1-39.

Irwin, J. J. (1994) Proneness to Dissociation and Traumatic Childhood Events. *Journal of Nervous and Mental Disease, 182,* 456-460.

Kapferer, B. (1983). *A Celebration of Demons: Exorcism and the Aesthetics of Healing in Sri Lanka*. Bloomington: Indiana University Press.

Kent, A., & Waller, G. (1998). The Impact of Emotional Abuse: An Extension of the Child Abuse and Trauma Scale. *Child Abuse & Neglect, 22,* 393-399.

Kleinman, A., & Sung, L. H. (1979). Why Do Indigenous Practitioners Successfully Heal? A Follow-up Study of Indigenous Practice in Taiwan. *Social Science and Medicine, 13B*, 7-26.

Kleinman, A. (1980). *Patients and Healers in the Context of Culture: An Exploration of the Borderland Between Anthropology, Medicine, and Psychiatry*, Berkeley, CA: University of California Press.

Krippner, S. (1994). Cross-Cultural Treatment Perspectives on Dissociation Disorders, In S. J. Lynn and J. W. Rhue (Eds.) *Dissociation: Clinical and Theoretical Perspectives* (pp. 338-361). New York: Guilford Press.

Krippner, S. & Colodzin, B. (1989). Multi-Cultural Methods of Treating Vietnam Veterans with Post-Traumatic Stress Disorder. *International Journal of Psychosomatics, 36*, 79-85.

Krippner, S. & Welch, P. (1992*). Spiritual Dimensions of Healing: From Native Shamanism to Contemporary Health Care.* New York: Irvington.

Kakarec, K. & Persinger. M. A. (1990). Electroencephalographic Validation of a Temporal Lobe Signs Inventory in a Normal Population, *Journal of Research in Personality, 24*, 323-337.

Kumar, V. K. & Pekala, R. J. (2001). Relation of Hypnosis-Specific Attitudes and Behaviors to Paranormal Beliefs and Experiences: A Technical Review, In James Houran and Rense Lange (Eds.), *Hauntings and Poltergeists: Multidisciplinary Perspectives* (pp. 260-279). Jefferson, NC: McFarland and Company.

Laderman, C. (1991). Taming the Wind of Desire: Psychology, Medicine, and Aesthetics in Malay Shamanistic Performance. Berkeley, CA: University of California Press.

Laderman, C. (1987). The Ambiguity of Symbols in the Structure of Healing, *Social Science and Medicine, 24*, 293-301.

Lambo, T. A. (1974). Psychotherapy in Africa. Psychotherapy and Psychosomatics, 24, 311- 326.

Lewis, I. M. (1971). Ecstatic Religion: An Anthropological Study of Spirit Possession and Shamanism. Middlesex, UK: Penguin Books.

Lewis-Williams, J. D., & Dowson, T. A. (1988). The Signs of All Times: Entoptic Phenomena in Upper Paleolithic Art. *Current Anthropology, 29*, 201-245.

Lynn, S. J. & Rhue, J. W., eds.(1994). *Dissociation: Clinical and Theoretical Perspectives*. New York: Guilford Press.

Ludwig, A. (1966). Altered States of Consciousness. *Achieves of General Psychiatry, 15*, 225-234.

Lumsden, C. L. & Wilson, E. O. (1983). *Promethean Fire: Reflections on the Origin of Mind.* Cambridge, MA: Harvard University Press.

Mac Hovec, F. J. (1975). Hypnosis Before Mesmer. *American Journal of Clinical Hypnosis, 17*, 215-220.

Margolis, C. G., Domangue, B. B., Ehleben, C.. & Shrier, L. (1983). Hypnosis in the Early Treatment of Burns: A Pilot Study. *American Journal of Clinical Hypnosis, 26*, 9-15.

McClenon, J. (1984). *Deviant Science: The Case of Parapsychology*. Philadelphia, PA: University of Pennsylvania Press.

McClenon, J. (1994). *Wondrous Events: Foundations of Religious Belief*. Philadelphia, PA: University of Pennsylvania Press.

McClenon, J. (1997a). Shamanic Healing, Human Evolution, and the Origin of Religion, *Journal for the Scientific Study of Religion, 36*, 345-354.

McClenon, J. (1997b). Spiritual Healing and Folklore Research: Evaluating the Hypnosis/Placebo Theory. *Alternative Therapies in Health and Medicine 3*, 61-66.

McClenon, J. (2000). Content Analysis of an Anomalous Memorate Collection: Testing Hypotheses Regarding Universal Features. *Sociology of Religion, 61*, 155-169.

McClenon, J. (2001). The Sociological Investigation of Haunting Cases. In J. Houran and R. Lange (Eds.), *Hauntings and Poltergeists: Multidisciplinary Perspectives* (pp. 62-81). Jefferson, NC: McFarland and Co.

McClenon, J. (2002a). *Wondrous Healing: Shamanism, Human Evolution and the Origin of Religion.* DeKalb, IL: Northern Illinois University Press.

McClenon, J. (2002b). Content Analysis of an Anomalous Experience Collection: Evaluating Evolutionary Perspectives. *Journal of Parapsychology, 66*, 291-316.

McClenon, J., Roig, M., Smith, M.D., & Ferrier, G. (2003). The Coverage of Parapsychology in Introductory Psychology Textbooks: 1990-2002. *Journal of Parapsychology* 67, 167-179.

McClenon, J. (in press). How Shamanism Began: Human Evolution, Dissociation, and Anomalous Experience, In J. Houran and R. Lange (Eds.), *From Shamans to Scientists: Humanity's Search for Spirits*. Jefferson, NC: McFarland

and Co.

McGuffin, P. , Riley, B., & Plomin, R. (2001). Genomics and Behavior: Toward Behavioral Genomics, *Science*, *291* (16 Feb.), 1232-1249.

Millar, M. G. & Millar, K. U. (1996). The Effects of Direct and Indirect Experience on Affective and Cognitive Responses and the Attitude-Behavior Relation. *Journal of Experimental Social Psychology, 32*, 561-579.

Moerman, D. E. (1979). Anthropology of Symbolic Healing. *Current Anthropology, 20*, 59-80.

Moore, L. E. & Kaplan, J. Z. (1983). Hypnotically Accelerated Wound Healing. *American Journal of Clinical Hypnosis, 26*, 16-19.

Morgan, A. H. (1973). The Heritability of Hypnotic Susceptibility in Twins. *Journal of Abnormal and Social Psychology, 82*, 55-61.

Morgan, A. H., Hilgard, E. R., & Davert, E. C. (1970). The Heritability of Hypnotic Susceptibility of Twins: A Preliminary Report. *Behavioral Genetics, 1*: 213-24.

Owen, I. M. & Sparrow, M. (1976). *Conjuring Up Philip: An Adventure in Psychokinesis*. New York: Harper and Row.

Palmer, J[ack] A., & Palmer, L. K. (2002). *Evolutionary Psychology: The Ultimate Origins of Human Behavior*. Boston: Allyn and Bacon.

Palmer, J[ohn]. A. 1979. A Community Mail Survey of Psychic Experiences. *Journal of the American Society for Psychical Research, 73*, 221-251.

Pekala, R. J., Kumar, V. K., & Marcano, G.. (1995). Anomalous/Paranormal Experience, Hypnotic Susceptibility, and Dissociation. *Journal of the American Society for Psychical Research, 89*, 313-332.

Pekala, R. J., Kumar, V. K., & Cummings, J. 1992. Types of High Hypnotically-Susceptible Individuals and Reported Attitudes and Experiences of the Paranormal and Anomalous. *Journal of the American Society for Psychical Research, 86*, 135-150

Persinger, M. A. (1984a). People Who Report Religious Experiences May Display Enhanced Temporal Lobe Signs. *Perceptual and Motor Skills, 58*, 963-975.

Persinger, M. A. (1984b). Propensity to Report Paranormal Experiences is Correlated with Temporal Lobe Signs. *Perceptual and Motor Skills, 59*, 583-586.

Persinger, M. A., & Makarec, K. (1987). Temporal Lobe Epileptic Signs and Correlative Behaviors Displayed by Normal Populations, *Journal of General Psychology, 114*, 179-195.

Persinger, M. A. & Makarec, K. (1993). Complex Partial Epileptic Signs as a Continuum from Normals to Epileptics: Normative Data and Clinical Populations. *Journal of Clinical Psychology, 49*, 33-45.

Persinger, M. A. & Valliant, P. M. (1985). Temporal Lobe Signs and Reports of Subjective Paranormal Experiences in Normal Population: A Replication. *Perceptual and Motor Skills, 60*, 903-909.

Putnam, F. W. (1985) Dissociation as a Response to Extreme Trauma. In R. P. Kluft (Ed.*), Childhood Antecedents of Multiple Personality* (pp. 66-97). Washington, DC: American Psychiatric Press.

Putnam, F. W. (1991). Dissociative Disorders in Children and Adolescents: A Developmental Perspective. *Psychiatric Clinics of North America, 14*, 519-531.

Richards, D. (1991). A Study of the Correlations Between Subjective Psychic Experiences and Dissociative Experiences. *Dissociation, 4*, 83-91.

Roig, M., Icochea, H., & Cuzzucoli, A. (1991). Coverage of Parapsychology in Introductory Psychology Textbooks. *Teaching of Psychology, 18*, 157-160

Ross, C. A. (1997). *Dissociative Identity Disorder: Diagnosis, Clinical Features, and Treatment of Multiple Personality*, New York: John Wiley & Sons.

Ross, C. A., Ryan, L., Voight, H. & Eide, L. (1991). High and Low Dissociators in a College Student Population. *Dissociation, 4*, 147-151.

Ross, C. A. & Joshi, S. (1992). Paranormal Experiences in the General Population. *Journal of Nervous and Mental Disease, 180,* 357-361.

Sanders, B., & Becker-Lausen, E. (1995). The Measurement of Psychological Maltreatment: Early Data on the Child Abuse and Trauma Scale. *Child Abuse & Neglect, 19*, 315-323.

Sharon, D. (1978). *Wizard of the Four Winds: A Shaman's Story*, New York: Free Press.

Spiegel, D. (1986). Dissociating Damage. *American Journal of Clinical Hypnosis, 29,* 123-131.

Swatos, W. H., Jr., & Gissurarson, L. R. (1997). *Icelandic Spiritualism: Mediumship and Modernity in Iceland,* New Brunswick, NJ: Transaction Publishers.

Targ, E., Schlitz, M., & Irwin, H. J. (2000). In E. Cardeña, S. J. Lynn, and S. Krippner (Eds.), *Varieties of Anomalous Experience: Examining the Scientific Evidence* (pp. 219-252). Washington, DC: American Psychological Association.

Thalbourne, M. A., Bartemucci, L., Delin, P. S., Fox, B., & Nofi, O. (1997). Transliminality: Its Nature and Correlates. *Journal of the American Society for Psychical Research, 91,* 305-331.

Thalbourne, M. A., & Delin, P. S. (1994). A Common Thread Underlying Belief in the Paranormal, Creative Personality, Mystical Experience and Psychopathology. *Journal of Parapsychology, 58,* 3-38.

Thalbourne, M. A. & Delin, P. S. (1999). Transliminality: Its Relation to Dream Life, Religiosity, and Mystical Experience. *International Journal for the Psychology of Religion, 9,* 35-43.

Ullman, M. (1947). Herpes Simplex and Second Degree Burn Induced Under Hypnosis. *American Journal of Psychiatry, 103,* 823-830.

Utts, J. (2001). *Annual Report for 2001*, Durham, NC: Parapsychological Association.

Varvoglis, M. P. (1999). *Annual Report for 1999*, Durham, NC: Parapsychological Association.

Vogel, V. (1970). *American Indian Medicine*. New York: Ballantine.

Völgyesi, F. A. (1966). *Hypnosis on Man and Animals*, 2nd ed., revised in collaboration with G. Klumbies. Baltimore: Williams and Wilkins.

Waller, N. G., Kojetin, B. A. Bouchard, Jr., T. J., Lykken, D. T., & Tellegen, A. (1990). Genetic and Environmental Influence on Religious Interests, Attitudes, and Values: A Study of Twins Reared Apart and Together. *Psychological Science, 1*, 138-42.

Wickramasekera, I. E. (1987). Risk Factors Leading to Chronic Stress-Related Symptoms. *Advances: Journal of the Institute for the Advancement of Health, 4,* 9-35.

Wickramasekera, I. E. (1988). Clinical Behavioral Medicine: Some Concepts and Procedures. New York: Plenum Press.

Winkelman, M. (1992). Shamans, Priests and Witches: A Cross-Cultural Study of Magico-Religious Practitioners, Tempe, AZ: Arizona State University Anthropological Research Papers No. 44.

Winkelman, M. (2000). Shamanism: The Neural Ecology of Consciousness and Healing. Westport, CT: Bergin and Garvey.

Young, D. & Goulet, J. (Eds.). (1994). Being Changed by Cross-Cultural Experience: The Anthropology of Extraordinary Experience. Peterborough, Ontario: Broadview.

Address for correspondence: James McClenon, Department of Social Sciences, Elizabeth City State University, Elizabeth City, NC 27909, USA.
E-mail: jmmcclenon@mail.ecsu.edu

STRANGE PEOPLE – RARE EVENTS

Peter Mulacz

Austrian Society for Parapsychology and Border Areas of Science, Vienna

ABSTRACT

A number of extremely small groups within the population that are in the possession of distinct 'gifts' or 'abilities' such as absolute pitch, synesthesia, and mental calculations skills, are identified and placed in a parallel with the likewise small group of 'sensitives' and 'mediums' in parapsychology. The widely discussed question whether 'psi' is 'democratically' distributed is touched upon as it appears that these other abilities are definitely not equally distributed. (This is a touchy subject in our society: whether or not all humans are in a certain sense equal or whether there are some differences.) Some parallels (and differentiations, too) between the two groups – absolute pitch, synesthesia, and mental calculation skills on one hand and 'sensitives' and 'mediums' on the other – are examined, e. g. the fact that these properties can hardly be trained – either they are present or they are not. Absolute pitch and mental calculations are 'composite' functions, in both cases the ability to memorize plays an important role. The question raises whether 'psi' is perhaps also a cluster of different abilities, apart from the problem that the entire notion of 'psi' as an 'ability' is highly questionable. Another question is the consistency of performance, be it the performance of one individual at different times (intrasubjective) or be it within several individuals (intersubjective) as in the case of comparative studies of sensitives and mediums. The emphasis of the paper, however, is on the fact that mainstream psychology and physiology carry out research into the particular abilities of very special persons who are clearly different from the vast majority and who form an extremely tiny minority of the population (in the case of the Mathematical Wizards just a handful individuals all over the world at any one time). From the point of philosophy of science this research may serve as a precedent for the legitimacy of research into the likewise rare 'gifted' people called sensitives or mediums. Thus their small sample size cannot be used as an argument against this kind of research. The lesser the number of individuals on one hand, the more outstanding and unique their performance on the other investigations will progressively shift of from the 'nomothetic' to the 'idiographic' pole.

Besides these 'strange' people the situation of rare or even unique events is discussed. Here again the idiographic and the nomothetic approach are balanced against one another, viewing upon the 'rare events' once more from the point of philosophy of science. Thus, without going into details of methodological issues it is argued that research into these fields, too, has some precedence in mainstream science. Not only is such research legitimate, moreover, any challenges to it (both from the skeptics' camp and from inside the parapsychological community) based on the sole ground of the rareness of cases must be rejected with reference to the outlined precedents in mainstream science.

INTRODUCTION

The shift to the 'Rhinean paradigm' bears a double importance: a theoretical one for deeper understanding of psi and a 'practical' one for the academic integration of parapsychology. It appears that 'strange' people with ostensibly 'strange' gifts and that 'rare' events like the ones the pre-Rhinean school of parapsychology was dealing with do not conform with the framework of scientific thinking shaped to the Galilean model of the (ideal) experiment that can be replicated at any place and any time irrespective of the persons involved. Thus the Rhinean approach was eagerly embraced for two reasons, (1) 'psi' was understood as a 'democratically' distributed ability[1, 2], and (2) Rhine's methodology came close to the ideal of an arbitrarily replicable experiment.

[1] From the point of history of science, it is worthwhile to consider these developments embedded in general history, i. e. the notion of democracy during various eras. Prior to World War I, most of the leading European powers were monarchies. Though at least the British, the Austro-Hungarian, and the German Empire were constitutional ones still aristocracy played a major role. Some time after the downfall of the Central Powers, fascism came into fashion almost all over Europe (Italy, Germany, Austria, Spain, Portugal, and Finland) doing away with democracy in favor of ideas of leadership and strength. (Mulacz, 2001) These are necessarily elitist and not egalitarian ideas. Eventually, when democratic regimes were installed in these countries, this shift in political ideas went parallel

Yet – is that really so? How shall we – from the point of philosophy of science – deal with people who claim to be specifically gifted, way beyond the average person? How shall we deal with rare events? Do they necessarily drop out of the framework of science?

Could it be that the Rhinean approach, successful though it has been, brought along the side effect that we narrowed our vision and thereby lost the view on the more promising field of specific psychic abilities?

However, things have swung back since. The extensive work in the field of Remote Viewing (Puthoff, 1996; Targ, 1996; May, 1996, etc.) has been carried out with particularly gifted subjects. Apparently, it is not so difficult to find such (Utts, 1996). Nonetheless this approach meets strong resistance from mainstream science. So it is worthwhile to look into that very mainstream science in order to see whether there are any areas in which particularly gifted persons are the subjects. If there are indeed such areas they may serve as a precedent that an approach based on rare particular abilities of only few individuals is not *per se* out of the scope of scientific thinking.

A somehow related question is how to deal with rare events (or even unique cases) from the point of philosophy of science.

STRANGE PEOPLE

Fields studying particularly gifted individuals

In my opinion there are – at least – four well established fields of that type[3]:

- Absolute pitch
- Synesthesia
- Mathematical Wizards
- Extraordinary Creativity

I will dwell only on the first three of them as it is difficult to find a measure for creativity just by the achievements – who was greater a genius, Michelangelo or Mozart? Clearly this would rather point in the direction of an idiographic than a nomothetic approach (idiographic = individualizing investigations such as clinical studies, case studies, historical investigations, as opposed to nomothetic research intending to find or establish 'natural laws').

Even in an experimental setting, the outcome would not likely be easy to judge (kind of parallel to free response tests in ESP research).

In the three other ones of the above fields there are some impressive results both in field studies and in experiments that merit a closer look.

Absolute Pitch

Since the pioneering days of German psychologist-physiologist Stumpf (Stumpf, 1883) experimental research on the phenomenon of absolute pitch is carried out. Until this day there are some 400 publications on this topic with every year half a dozen new ones being added. Tests mostly follow a pattern of perception of a certain sound, identification i.e. correlation to an acoustic memory, and announcement of that

with shifting from a parapsychology based on experiences with sensitives and mediums to the new style in parapsychology based on Rhine's work. Of course there have been good reasons for this ready acceptance, e.g. methodological issues, yet the question remains whether the apparent parallel with ideological changes was just by coincidence or perhaps deeper-rooted, of course subconsciously and Zeitgeist-related. I should add that my remark refers to just one aspect of the reception of Rhine's work particularly in post-war Europe, not to Rhine's own starting point.

[2] The issue of the distribution of psi as such has a history of long and partly controversial discussions in parapsychology. A relevant overview of the state of affairs is provided by Jürgen Keil's survey (Keil 1979).

[3] I am referring only to those fields where performances like the ones mentioned in this chapter are being accomplished by otherwise 'normal' people (regardless how to define 'normal'), thus their achievements come additionally (a surplus), as opposed to such persons where super-achievements in one field are balanced against serious shortcomings in others like in the well known cases of 'idiot savants'.

assignment. The point is the 'internal reference' or the long-term storage in the memory whereas the vast majority of people (including the majority of musicians) avail only of a short-term acoustic memory (relative pitch, which in its essence is the mere identification of intervals, whereby the memory would last only for seconds rather than minutes). The term 'absolute pitch' is used provided the identification rate is high (close to 100%) and its accomplishment is robust. Though the hit rate is higher with musicians than with average people yet there is a clear distinction between what can be achieved by even musical people and by those with absolute pitch. Depending on the level of the test, a sample of average people including the subset of musicians would score some 10–40%, whereas the absolute pitch endowed individuals would score 70–100% (with hardly someone in between these two groups) being able to process more than double the amount of information.

Whether this ability is inherited or acquired cannot be decided upon for sure. Also the question remains open whether the absolute pitch is the result of a number of several variables (as it is the fact with most cognitive abilities) or one single ability on its own. Efforts of training this 'ability' did not show significant results. Experiments focused on 'disturbing' that internal reference by other stimuli did not work either.

Estimates on the size of the group of persons with absolute pitch in the population vary, however, generally accepted is a value in the range of the fraction of one percent of the population.

Hence we encounter there a certain minority of people – a distinct group of persons – who appear to possess an 'ability' the vast majority of the population is lacking, an ability that appears not to be Gaussian or 'democratically' distributed. Individuals with this rare and strange ability are being scientifically studied since some 120 years. This may well serve as a precedent supporting the legitimacy of the study of equally rare specifically gifted persons (sensitives, mediums) in parapsychology.

Synesthesia

Synesthesia is an involuntary experience[4] of a cross-modal association which means the 'blending of senses', i. e., when one quality is being stimulated like smelling or hearing, another quality is simultaneously perceived like seeing colors or geometrical shapes. The most frequent one is "coloured hearing" ("audition coloreé") whereby typically sound, music, voices and characters/figures that are read aloud lead to the perception of moving colors and shapes that are being projected either into the outer world or into the interior of the head. On a kind of 'internal video screen' without sharp spatial delineation colored structures appear frequently, spheres or longish 3-dimensional entities with characteristic surfaces, e. g. velvety or glittering or glass-like or metallic. With 'genuine synesthetes', the character of these surfaces correlates directly to the perceived acoustic stimuli. Synesthesia appears to be more frequent with females and with non-right-handers, and it is a family trait. Yet some famous synesthetes are male: Russian composer Alexander Scriabin (1872-1915) and painter Vassily Kandinsky (1866-1944) who both wrote about their experiences. Some text portions of Isaac Newton and of Johann Wolfgang Goethe indicate that such experiences were not unfamiliar to them. Some artists and other creative people utilize their synesthetic experience in their productions.

Estimates of the percentage of synesthetes within the population vary a lot, between 1 in 2 000 und 1 in 1 000 000. A generally accepted average is 0.04 ‰ (Cytowic, 1989, 1993).

Again we encounter a small fraction of the population (much smaller than the formerly discussed people bestowed with absolute pitch) in the possession a well-described "ability" peculiar to them and not present with the majority. Synesthetes are clearly different from the average people, however, other than the people with absolute pitch who have a better memory and a wider information processing bandwidth it is not quite clear whether synesthesia is a plus, an advantage, over others, or just a deviation. Yet for our purpose – looking for parallels to studying particularly gifted subjects in mainstream psychology/ physiology – it will do

[4] This of course leads to philosophical questions such a 'What is real?', 'Real – to whom? To the experiencer or to everyone?' 'Do we all live in the same reality', and, ultimately, uttered in a provocative way, 'What is really real?'. (From a psychological viewpoint rather than from the philosophical one Watzlawick tackles the same question in a book the programmatic title of which reads 'How Real is Real?'.)

to note that the phenomenon of synesthesia falls into this category and that the percentage of people concerned is extremely small.

Mathematical Wizards

There are plenty of reports on mathematical prodigies in the literature – it seems every period has their own ones. Some twenty years ago I had the opportunity to attend a demonstration Willem (Wim) Klein gave in the Institute of Physics at the University of Vienna. Klein, a Dutchman who at the time of this demonstration was already retired used to be employed by CERN (the European Nuclear Research Center) in Geneva, Switzerland, to serve as a "human computer".

Fig. 1 Wim Klein cartoon

The demonstration of his calculation skills was absolutely fascinating. Not only did he carry out various mathematical operations (multiplication, division, root extraction, squaring, cubing, and even higher powers) with very large figures, he also gave explanations how he did it. The precondition, of course, is the ability to memorize those multi-digit figures, an ability exceeding the possibilities of the average person by several orders of magnitude. There is nothing paranormal or mystical or intuitive in his computing, as he emphasized, he simply calculates mentally in a similar way one might do with pencil and paper, he looks whether a figure would be a prime number, also, by which divisor that figure could be divided, he fractionizes large figures into smaller ones, carries out calculations with those and adds them up mentally, etc. By doing so he needs to keep a number of different partial results in his memory at any time without mixing them up with one another. He was able to determine the number of digits of a large figure just by looking at it, without counting them – amazing. Unfortunately Willem Klein was killed in a car accident in his native Netherlands not too long after this demonstration.

A search on the web reveals that there are number of contemporary mathematical wizards who are well known – primarily Gert Mittring from Germany. A few examples of his achievements:

- Memorizing of 22 decimal figures within in 4 seconds as well as 30 binary figures within 3 seconds
- Extraction of the der square root of a six-digit figure within 44.7 seconds by mental arithmetic
- Extraction of the 13th root of a 100-digit figure within 39.0 seconds by mental arithmetic

- Calendar-Memory (error-free identification of the days of the week at 20 randomized days of the century) within 38.2 seconds as well as to 20 dates of the years 1600-2100 within one minute

Fig. 2 Dr. Gert Mittring (photo taken from the web)

These mathematical prodigies even carry out their own international competitions, like Olympics. Find below a few examples of their results indicating the time needed to fulfill the task and the date and place of the competition. Currently, Gert Mittring holds 14 world championships in mental calculation[5] and enjoys several entries in the 'Guinness Book of Records'. Root extraction is his 'specialty'. He has published a book 'What happens in us when we calculate?' (German language only) based not only on his introspection into his experiences but also on questioning others.

Mental calculation of the square root of a 6-digit figure:

- 64 seconds M. Hari Prasad (India) 10/30/1999, Bangalore
- 52 seconds Gert Mittring (Germany) 2000, Linz
- 45 seconds Gert Mittring (Germany) 07/07/2000, Flensburg

Mental calculation of the 13th root of a 100-digit figure
(whereby the 100-digit figure is chosen that way that it is x to the power of 13):

- 322 seconds Willem Klein (Netherlands) 09/19/1975, Amsterdam
- 231 seconds Willem Klein (Netherlands) 11/08/1978, Stockholm
- 129 seconds Willem Klein (Netherlands) 05/06/1980, London
- 116 seconds Willem Klein (Netherlands) 11/13/1980
- 89 seconds Willem Klein (Netherlands) 04/07/1981, Tsukuba
- 39 seconds Gert Mittring (Germany)

[5] Funnily enough he was not a good student in mathematics during his years in school.

Mental calculation of the 23rd root of a 200-digit figure
(whereby the 200-digit figure is chosen that way that it is x to the power of 23):

- 50 seconds Shakuntala Devi (India) 1977

Mental calculation of the 73rd root of a 500-digit figure
(whereby the 500-digit figure is chosen that way that it is x to the power of 73):

- 2:43 minutes Willem Klein (Netherlands) 08/27/1976, Geneva
- 2:09 minutes Willem Klein (Netherlands)
- 1:43 minutes Willem Klein (Netherlands) 11/22/1983, Hamburg

As mentioned before, the precondition for the performance of such amazing mental calculation skills is the ability to memorize the large figures involved. Regardless whether we split the overall ability of mental arithmetic into two (memory and calculation) or consider it as whole, in any case we encounter an extremely small but well-defined group of individuals within the population. This group (at least the winners and those coming in second) consists of just a handful of individuals in Europe or even world-wide. These few individuals avail of an ability (or of a set of abilities) proprietary to them and alien to the population at large. They show consistent results in their performances, i. e. their performance is replicable. Moreover, not only is it replicable within themselves (what I call intra-subjectively, e. g. at different points in time), it also replicable amongst them (inter-subjectively, e. g. all of them achieve somehow similar results). By virtue of this intersubjective replicability it is justified to state they form a certain group of specifically gifted people, a (small) subset of the population that is different from the majority. Likewise the fact that they are considered as a group (and not as individuals) corresponds to the nomothetic approach as opposed to an idiographic one if focusing on one individual only.

Summary

So far, we have considered three different groups of people with very rare and specific abilities the vast majority of the population does not avail of – in other words, abilities that are apparently not 'democratically' distributed and that show no distinct gradient, i. e. the ability is either present (which refers to the individuals within these groups) or not (which refers to the population at large). The size of these groups is in any case very small, yet there are huge differences size wise between the groups:

- Absolute Pitch — a fraction of a percent of the general population
- Synesthetes — between 0.5‰ and 0.001‰ of the general population (estimate of 0.04‰)
- Mathematical Wizards — a mere handful of individuals world-wide

Individuals belonging to each of these groups are the subjects of scientific investigations and research.

This type of research into small and smallest groups tagged as particularly gifted ones is a precedent for the (comparative) investigation of gifted persons within parapsychology, i. e. sensitives and mediums. From the point of philosophy of science these two lines of research are running perfectly parallel, the research in the above three groups (absolute pitch, synesthetes, and mathematical wizards) and the research in sensitives and mediums are equally justifiable. Research in the particular abilities of e. g. Wim Klein and Gert Mittring is a parallel to the research in the particular abilities of e. g. Stephan Ossowiecki and Gerard Croiset.

Having emphasized these parallels from the point of philosophy of science, I should mention the differences between the group of persons with Absolute Pitch, Synesthetes and Wizards on one hand and the group of mediums and sensitives in parapsychology on the other. There might be two problem areas, on is a rather formal one (the question of definition – we all know what mediums and sensitives are, but supplying a generally accepted definition, drawing a generally accepted dividing line between the "normal" and what is beyond the normal is much more difficult than in the case of Synesthetes and Wizards). The other one is a methodological one. It appears that the first group (Absolute Pitch, Synesthetes, and Wizards) are a much more homogenous group than the other, the mediums and sensitives, thus it is much easier to research and describe their abilities and compare it to the average person. The mediums and sensitives are no coherent group and many of them would deserve individual studies, i. e. the nomothetic approach (as opposed to a standardized test such as a hearing test). When going slightly more into detail, we could perhaps distinguish between such mediums and sensitives who might be grouped and compared to one another, e. g. Maria Reyes de Z., Gerard Croiset, and Arthur Orlop in one group, or the contemporary RVers in another, or some "physical mediums" such as D. D. Home, Eusapia Paladino and the Schneider brothers. Nonetheless the achievements of mediums within one of these groups are much more at variance than the achievements of Synesthetes or Wizards. On the other hand, the "better" (and sometimes even more bizarre) the performance of a medium, the more he or she goes beyond the "normal" and the "average", the lesser the importance of the question of a proper definition. Gerd Mittring's mental calculation skills are far beyond the "normal" or the "average", and so is Rudi Schneider's ability of moving objects in his vicinity without using "normal" motor means.

Another parallel between the 'gifted people' under review and the mediums/sensitives (and perhaps another difference, too) need to be considered. For persons with Absolute Pitch, we learned that efforts of training this 'ability' did not show significant results which will remind parapsychologists about the situation with sensitives and RVers. Also, we learned that (as the experts in the field say) the question remains open whether the absolute pitch is the result of a number of several variables (as it is the fact with most cognitive abilities) or one single ability on its own – yet there is a strong impression that there are at least two variables involved, namely hearing and memorizing. Memory is also a variable – rather *the* variable – in the case of the Mathematical Wizards. Thus is might be justifiable to say that in the various cases of 'gifted people' the results of their amazing productions are based on a cluster of different abilities. How is this in the case of sensitives and mediums? It might perhaps be that we encounter a difference between the 'psi' ability of sensitives/mediums and 'ordinary psi' as measured in small deviations from chance in mass experiments that is similar to the difference between the hearing ability of people with absolute pitch and 'ordinary' people or the mental calculations of Mathematical Wizards vs. the ones of others.

Do we know enough about the nature of 'psi' to decide this question? Moreover, is it correct to talk in the context of mediums and sensitives about an 'ability'? Usually the notion of an ability goes together with the possibility of using that ability at will. If Gert Mittring decides to take part in a mental calculation competition he simply does so, and the outcome is (talking about orders of magnitude) predictable. This is generally not the case with mediums and sensitives. Often they must be in a certain mood. At times the observer gets the impression things 'happen' to the medium, not the medium is 'doing' something. It is difficult to find a conceptional framework encompassing the properties of all the known cases, even if splitting them into different groups (such as psi-gamma, psi-kappa). Anyway, this goes into details of parapsychology[6] whereas I had intended to content myself with viewing upon this issue only from the epistemological point of view or from philosophy of science.

Let me conclude this chapter by summarizing that within the population at large there are several very small groups identified that are in possession (or claiming to be so) of abilities the majority of the population does not avail of. Some of this groups are being well researched, some are less (e. g. the mediums and sensitives, for various reasons). They do not only have the same right to be investigated with all scientific scrutiny, no argument can be raised against this on the sole ground that this is an extremely small group or even a few isolated individuals. Such argument needs to be rejected, whether it originates in the camps of the skeptics or within the parapsychological community[7] itself, and reference to the ample research in the fields described above is a good criterion on which this refutation may be based.

[6] I was encouraged to add a few thoughts in this direction by my (anonymous) referees to whom I feel indebted.

[7] It might be recalled that John Beloff distinguishes between two camps, the 'maximalists' and the 'minimalists'. Pursuing that matter, relevant though it is, would lead even further away from my main points.

RARE EVENTS

General

A related yet different topic is how to deal with rare (or even unique) events. As an example of such within parapsychology we might think either of various spontaneous cases or of experiments that found no (or only very limited) replication, e. g. Zöllner's experiments with Henry Slade in 1878 achieving knots being knotted into 'continuous' cords or leather stripes. There were several instances of this experiment that can be viewed upon as different stages within the process of improvement and refinement of the one basic experiment. Certainly they show Slade's consistent ability to achieve basically the same results, i. e. a certain amount of replicability, however, this (conceptual) replicability was achieved by the same persons (experimenter and psychic alike) under partly unchanged conditions and during a short timeframe. Thus it appears justified to consider this series of experiments as a whole, like one event. Later replication attempts did not bear fruit, at least those under controlled conditions (e. g. the wooden rings with 'Margery' = Mrs. Crandon). Thus it would appear safe to put Zöllner's experiments with Slade into the category of 'rare events'.

Again in order to answer the question how to deal scientifically with such rare events (to take them into serious consideration despite their rareness or simply to dismiss them at all due to their rareness) I am looking for precedents of 'rare events' in mainstream science.

Supernovae

Back in 1572, famous Danish astronomer Tycho de Brahe made an observation of a supernova in Cassiopeia (literally: novus, -a, -um = new, sc. star, i. e. "new star").

In 1604, both Kepler and Galileo (as well as Chinese and Korean astronomers) observed and recorded a "new star" in Serpens (another supernova), in 1610 Galileo published his "Sidereus Nuncius" (Starry Messenger) in Venice.

Though the interpretation has changed since (supernovae are by no means "new stars") the mere facts of these observations have not been rejected, neither at their time nor ever since. The second case has luckily been observed by more than one individual simultaneously which adds to its credibility but also Tycho's sole observation had been accepted in its own right.

Coelacanth

Near the end of 1938, the first living Coelacanth (an archaic fish later named 'Latimeria chalumnae Smith') was discovered off the Eastern coast of South Africa. The prominent ichthyologist J. L. B. Smith remarked he could not have been more surprised about this 'living fossil' as if he had encountered a living dinosaur in the street.

This first Coelacanth caught in 1938 led to the discovery of an entire population, between the mainland of Africa and Madagascar. For decades to come this was presumed to be the only Coelacanth population surviving to this day.

However, in 1998, another Coelacanth population was discovered off the island of Sulawesi, Indonesia, located some 10,000 km distant from the area of the population that has been discovered firstly.

So we have three steps: one individual animal, one isolated population, and, ultimately, more than one single population. Yet the credibility of the basic discovery has not been called into doubt even before these entire populations were found.

Summary

Comparing those two instances of 'rare events' there are aspects they have in common and others that are different. What is common to both is – besides their rareness by definition – that the very first observations in these cases came spontaneously and by surprise. What is different is the number of repetitions. In the case of supernovae, the repetitions are comparatively rare, whereas in the Coelacanth case entire populations have been discovered. When comparing those to case studies in parapsychology we find the same spectrum from unique, isolated cases to such that found later repetitions, either with the same individual or with others (intra-subjectively vs. inter-subjectively). As it has been the case with those 'strange people' discussed above, there is – from the viewpoint of philosophy of science –no difference in principle between the situation in mainstream science and in parapsychology, and the 'rare' cases in mainstream science may serve as precedents for the treatment of 'rare' parapsychological cases. Thus rare occurrences in parapsychology, provided they are well documented, ought not to be rejected on the sole ground of their rareness.

Discussion

The link between the 'strange people' and the 'rare events' is the notion of rareness as such as those 'strange people' are rarely found within the population. This very rareness is frequently viewed upon as being opposed to one basic aspect of science, the repeatability. Repeatability, however, makes a number of tacit assumptions, foremost that one that there are no variables undergoing a change between an initial and a replicated experiment unless the experimenter introduces such changes deliberately. In particular this means that variables are independent from the point in time the experiment or its replication is performed. Without entering the difficult question "what is time" a simple consideration of the 'time arrow' shows that no two moments are really the same, in other words, the notion of repeatability or replicability as such is highly questionable. Thus an element of the idiographic approach (i. e. honoring individual cases) inevitably enters the field of nomothetic (general law-related) science as well. From this viewpoint it is only a small step forward to extend the field around the idiographic pole.

Ever since the days of Ludwik Fleck and Thomas S. Kuhn philosophy of science looks into the history of science, too, in order to establish or rather identify guiding principles that have proven to be valid across considerable time spans. Thus it seems reasonable to look into the 'past practice' in fields well established in mainstream science for precedences applicable on parapsychology.

Many scientific studies into the particular abilities of individuals endowed with absolute pitch, synesthesia, or calculation skills have been carried out in the past. They may serve as precedents for studies of equally rare particularly 'gifted' individuals within parapsychology. In other words, charging parapsychology on the ground that the field is studying 'strange people' with 'rare' abilities is unfounded and clearly illegitimate. Luckily such kind of resistance and refutation is waning.

The astronomical and biological cases of the 'rare' events as described above have been accepted whereas equally 'rare' events in the field of parapsychology were met with fierce resistance, even from inside the field. The only reason for such different treatments may be the fact that the rare events in 'normal' science are viewed upon as being easier to accommodate within the framework of scientific knowledge. It is highly questionable whether this is really the case. The 'rare events' in the field of parapsychology can mostly be accommodated within the known framework of parapsychology itself however there are two requirements to do so: knowledge is needed as older cases should not be buried in oblivion but taken into account for comparison purposes, and, admittedly, a certain amount of courageousness is needed as well.

References

Barrington, M. R. (nd). *Ossowiecki*. Unpublished manuscript/personal communication.

Cytowic, R. E. (1989). *Synesthesia: A Union of the Senses.* New York: Springer.

Cytowic, R. E. (1993). *The Man Who Tasted Shapes: A Bizarre Medical Mystery Offers Revolutionary Insights into Reasoning, Emotion, and Consciousness.* New York: Putnam.

Fleck, L. (1979). *Genesis and Development of a Scientific Fact* (originally published in German, 1935) Chicago: University of Chicago.

Gregory, A. (1985). *The Strange Case of Rudi Schneider.* Metuchen, NJ: Scarecrow Press, Inc

Keil, H. H. J. (1979). The distribution of psi ability and psi performance. *European Journal of Parapsychology, 3,* 177-183.

Kopfrechnen: Wurzelziehen. <http://www.recordholders.org/de/list/roots.html>

Kümmel, A., & Steckiewicz, J. (2003). Leipzig 1877: Medienepistemologische Zugänge zu Karl Friedrich Zöllners Experimenten mit Henry Slade. In: Orte des Okkulten (Geppert, A. C. T. & Braidt, A. B., Eds.). *Österreichische Zeitschrift für Geschichtswissenschaften. 14,* 72-95.

Kuhn, Th. S. (1962). *The Structure of Scientific Revolutions.*

Lang, A. (1993). Das "absolute Gehör" oder Tonhöhengedächtnis. In: Bruhn, Oerter & Rösing (Eds.) *Musikpsychologie – ein Handbuch.* Reinbek: Rowohlt.

Lucadou, W. v. (1990). Was man nicht wiederholen kann – zum Problem der Replizierbarkeit. *ZfPuGdP, 32,* 212-230.

May, E. C. (1996). The American Institutes for Research Review of the Department of Defense's Star Gate Program: A Commentary. *Journal of Scientific Exploration. 10,* 89-110. Also in: *The Journal of Parapsychology*. 60, 3-23.

Mittring, G. (2001). *Was geht in uns vor, wenn wir rechnen? Erkenntnisse und Erfahrungen des Weltrekordlers im Kopfrechnen.* Marburg: Tectum.

Mulacz, P. (2001). Biology, Vitalism and the 'Golden Era' of Continental European Parapsychology. *The Parapsychological Association, 44th Annual Convention, Proceedings of Presented Papers*, NY: 191-204.

Puthoff, H. E. (1996). CIA-Initiated Remote Viewing Program At Stanford Research Institute. *Journal of Scientific Exploration. 10,* 63-76.

Stumpf, C. (1883). *Tonpsychologie.* Leipzig: Hirzel.

Targ, R. (1996). Remote Viewing At Stanford Research Institute In The 1970's: A Memoir. *Journal of Scientific Exploration. 10,* 77-88.

Thouless, R., & Wiesner, B. (1948). The psi process in normal and "paranormal" psychology. *Journal of Parapsychology, 12,* 192-212.

Utts, J. (1996). An Assessment Of The Evidence For Psychic Functioning. *Journal of Scientific Exploration. 10,* 3-30. Also in: *Journal of Parapsychology.* 59, 289-320.

Watzlawick, P. (1976). How Real is Real? New York: Random House.

Zöllner, J. C. F. (1878). *Wissenschaftliche Abhandlungen.* Leipzig: Staackmann. Translation (1881): *Transcendental Physics: An Account of Experimental Investigations.* Reprint: Kessinger.

Address for correspondence: Peter Mulacz,
Österreichische Gesellschaft für Parapsychologie und Grenzbereiche der Wissenschaften, c/o Institut für Ethnologie, Kultur- und Sozialanthropologie der Universität Wien,
Universitätsstraße 7, A-1010 Wien.
E-mail: peter.mulacz@aon.at

BODY IMAGE IN RESPONDENTS WITH AND WITHOUT PRIOR OUT-OF-BODY-EXPERIENCES

Craig D. Murray & Jezz Fox
Liverpool Hope University College

ABSTRACT

Although the body is an implicit aspect of out-of-body experiences (OBEs), to date bodily experience itself has been an under-researched area in OBE research. The present study builds upon the important insights of three psychological approaches to understanding OBEs, namely the psychological theories of Palmer (1978), Blackmore (1984), and Irwin (2000). Palmer (1978) presented a psychological theory of the OBE that centred on the body image of the experient. He argued that a reduction of proprioceptive information from the body resulted in changes to the 'body concept' which (in some instances) in turn 'triggered' the OBE. Similarly, Blackmore (1984) proposed that the brain builds up a model of reality based in large part on sensory input; when there is a breakdown in this sensory-input model, combined with other cognitive processes, an OBE occurs. Irwin (2000) provides a third theory for the occurrence of the OBE, namely that it is the result of the convergence of a number of dissociative factors. This includes high levels of 'absorption' as well as a simultaneous occurrence of dissociation from somatic input. These theories of the OBE are informative exceptions to the general lack of consideration given to bodily experience in the OBE research literature. However, they have a very narrow focus on perceptual aspects of this experience. For instance, Palmer's and Blackmore's theories posit changes in the 'body concept' or 'body image' where these terms would seem to refer to conscious or unconscious changes in the sensorial topography of the body, such as might occur with a lack of somatosensory input. Irwin's study goes a little further in that it considers not just a lack of sensory input but amplified sensory inputs. However, what is over looked in these studies is other ways of experiencing the body. In the present study we theorised that the daily bodily experiences of people with a prior OBE and those without this experience would differ along a number of dimensions. In order to test this theory a web-based questionnaire study was conducted. Of 243 respondents, 62 reported at least one prior OBE (45 females, 17 males, with a mean age of 29 years, SD = 10.9). The non-OBE sample consisted of 181 participants (126 females, 55 males, with a mean age of 29 years, SD = 10.7). Six scales on different aspects of bodily experience were administered. Respondents reporting a previous OBE were found to score significantly higher on measures of somatoform dissociation ($p<.001$), self-consciousness ($p=.035$), body dissatisfaction ($p<.001$), and lower on a measure of confidence in their physical self-presentation ($p=.02$) than respondents without a previous OBE. There were no significant differences between these groups concerning physical self-efficacy or perceived physical ability, objectified body consciousness, or social physique anxiety. The findings within this paper lend support to our contention that there are pre-existing differences in the body experiences of those people who do and do not have OBEs which are not accounted for in current psychological theories.

INTRODUCTION

The Out-of-Body Experience (OBE), in which the person who has an OBE has an experience in which the self and their body are phenomenologically separated, has been a topic of concern in the psychological sciences for well over 100 years (Alvarado, 1992). Alvarado (1992) notes that the key features of an OBE often include a sensation of floating, seeing one's own physical body from outside, and an experience of travel to a place remote from one's actual physical-body location. While the body is an implicit aspect of out-of-body experiences, as a particular topic of interest it is generally overlooked in the research literature. Three important exceptions to this are the psychological theories of OBEs advanced by Palmer (1978) and Blackmore (1984), and a recent study (in which a third theory of the OBE is advanced) by Irwin (2000) into the somatosensory experience of OBE experients.

Palmer (1978) presented a psychological theory of the OBE that centred on the body image of the experient. He argued that a reduction of proprioceptive information from the body resulted in changes to the ‘body concept’ which (in some instances) in turn ‘triggered’ the OBE. Drawing on Freudian theory, Palmer argued that the person’s change in body concept threatened their self-concept or sense of individual identity. As a result of this threat unconscious processes are activated in an attempt to re-establish the person’s sense of individual identity. An example given by Palmer (1978) is one of “deathbed experiences” in which he argued that an OBE is more likely to occur when the experient first has had a prolonged confinement to bed in a weakened condition (which is likely to encourage changes in body image) and (whether he/she accepts it intellectually or not) they have a psychological set of death which involves the separation of the soul from the body. With the re-establishment of the normal body concept the OBE ends. It is important to note that Palmer views changes in the body image as necessary but not sufficient for an OBE to occur; it is the Freudian ‘primary process’ attempt at reintegration which, according to Palmer, in some cases results in an OBE.

A second psychological theory of OBEs is that provided by Blackmore (1984), who proposed a cognitive explanation for the occurrence and phenomenology of OBEs. She argued that one of the key functions of the brain is the construction of models of the self within the environment. These may be short-term models (such as those created through perception) or long-term models (such as those formed through memory) and each may influence the other. One of these models that Blackmore (1984, p.203) identifies is that of the body image, which she states is the sum of “somatosensory information, visual, and other sensory input and memory.”

Rather than modeling what is ‘really’ there, Blackmore (1984) points out that we ‘build’ models of reality. Moreover, we are continually engaged in multiple model making processes: a model of the activity we are currently engaged in, a model of a remembered event, a model of an imagined state of affairs, and so on. Central to Blackmore’s theory is the proposal that at any given time only one model is taken to represent ‘reality’. For most people at most times this ‘reality’ model will be taken to be that which is most complex, stable and coherent. This model would usually be that generated in most part through sensory input. However, at times sensory input may be reduced, as in sensory isolation experiments, resulting in impoverished sensory models. This may result in an over reliance on information from memory in order to achieve stability. Therefore for an OBE to occur at least two things are needed: the failure of the somatosensory input-controlled model, and the substitution of an imagery-based one built-up from memory. The required incapacitation of the input model leads Blackmore (1984) to argue that her theory explains why such experiences as sensory deprivation, relaxation, illness, and certain drugs frequently occasion OBEs.

Blackmore (1984) herself outlines the similarity between her approach and that of Palmer (1978), insomuch that both posit that OBEs occur as the result of a change in the body concept or body image. Therefore Palmer (1978) and Blackmore (1984) would seem to have a shared understanding of what “body image” actually is, and that this relates to a particular sensorially obtained and cognitively represented conceptualisation: Palmer sees the proprioceptively derived body concept as being suppressed during the occurrence on an OBE, while Blackmore sees the loss of body image based on a sensory input model of the self in the environment as beginning a chain of events which result in an OBE.

While the above two psychological theories have a strong emphasis on the role of body image in the occurrence of OBEs, other research has paid little attention to the experience of the body in relation to OBEs. However, one recent exception to this is a study by Irwin (2000) who has examined the experience of somatic symptoms and OBEs. Drawing upon the work of Nijenhuis and colleagues (e.g. Nijenhuis, Spinhoven, van Dyke, van der Hart and Vanderlinden, 1996) into dissociative states, Irwin argues that OBEs are in part the result of somatoform dissociation in which there can be a ‘deficit symptom’ such as numbness in a part of the body, or ‘positive symptom’ in which psychosomatic pain or tics are experienced. Irwin’s rationale for studying somatoform dissociation in OBEs is that “at a phenomenological level the OBE appears to entail a dissociation between sensory processing of somatic (somaesthetic and kinaesthetic) events and the sense of self or identity” (Irwin, 2000). In his study involving 113 psychology students, with an OBE incidence rate of 38.9%, Irwin (2000) administered the Somatoform Dissociation Questionnaire. This was found to be the only predictor variable (from a logistic regression analysis which included participants’ data for dissociative experiences, absorption, gender and age) able to independently discriminate between people with and without a prior OBE, as well as the only independent variable which contributed significantly in predicting OBE frequency.

Irwin (2000) provides a third theory for the occurrence of the OBE, namely that it is the result of the convergence of a number of dissociative factors. This includes high levels of 'absorption' (a psychological state in which the person is in a high state of engrossment in experience), as well as a simultaneous occurrence of dissociation from somatic input. This theory includes a reformulation of Irwin's earlier theory (1985) based upon his findings that people with prior OBEs exhibit a high capacity for psychological absorption, while people with high levels of psychological absorption were more susceptible to experimentally induced OBEs. These changes are posited to undermine the socially conditioned assumption that the body is the container of the self, and as a result to promote the feeling that the person's consciousness is no longer in the spatial confines of the body. In turn, this abstract perception of a disembodied self is cognitively processed "into a passive, generalized somaesthetic image of a static floating self" (Irwin, 2000) and into an experience of an OBE. Through the process of synesthesia, and providing the experient has a basic visuospatial ability, the somaesthetic image may be translated into a visual image. Irwin (2000) argues that strong absorption during the above state is responsible for the experienced realism of the OBE.

Again, there are similarities between Irwin's (2000) OBE theory and those of Palmer (1978) and Blackmore (1984). A breakdown between sensory input and experience of this input is central to all three theories (although the nature of this 'breakdown' is different in Irwin's theory). We believe that Irwin's (2000) findings are indicative of a new line of inquiry in relation to OBEs. While both Palmer and Blackmore concur that a change in body image precipitates the occurrence of an OBE, Irwin's findings are suggestive of pre-existing, possibly long-term differences between the bodily experiences of people who have (or who are likely to have) OBEs and those who have not (and are unlikely to do so in the future). Prior research has found that people who report having had an OBE are more likely to have had multiple rather than single OBEs; in an analysis of 19 studies Alvarado (1986) found on average 67% of experients reported more than one OBE. Neither Palmer's nor Blackmore's theories attempt to explain why some people are more prone to OBEs than others are. In contrast, Irwin's (2000) empirical work suggests some people with certain bodily experience are more likely to have an OBE. Furthermore, rather than OBEs occurring only in the absence or reduction of sensory information, the association between somatoform dissociation (which includes items that measure 'positive symptoms' or the *amplification* (rather than just the reduction or absence) of some sensory experiences, such as pain) indicates that the relationship between bodily experience and OBEs is more complex than that outlined in Palmer's (1978) and Blackmore's (1984) theories.

We would argue that OBEs are more likely to occur in people who usually have a *weak* sense of embodiment, characterised by a generalised dissociation between their self and body which can be measured on a number of body experience dimensions. To clarify this further, we believe that Palmer's and Blackmore's theories posit body image change or disturbance as an immediate precursor to the occurrence of OBEs rather than as a qualitative difference in the everyday bodily experience of people who do and do not have OBEs. This is not to challenge the central premise of these theories, namely that a 'change' in body image precipitates OBEs, but rather to complement these theories by suggesting that out-of-body experients have pre-existing differences in their body experience compared to non-experients, and that these pre-existing differences are possibly exacerbated in the moments in which an OBE occurs.

We would add to our argument here that although the theories of Palmer, Blackmore and Irwin (including the latter's empirical work described earlier), are informative exceptions to the lack of consideration given to bodily experience in the OBE research literature, they have a very narrow focus on perceptual aspects of this experience. For instance, Palmer's and Blackmore's theories posit changes in the 'body concept' or 'body image' where these terms would seem to refer to conscious or unconscious changes in the sensorial topography of the body, such as might occur with a lack of somatosensory input. Irwin's study goes a little further in that it considers not just a lack of sensory input ('deficit' symptoms) but amplified sensory inputs (such as pain, or 'positive' symptoms). However, what is over looked in these studies is other ways of experiencing the body.

Outside of parapsychology, body image studies can be divided into research areas which address persons' perceptual experience of their bodies (such as body size estimation, and sometimes referred to as the 'body percept'), their own subjective feelings towards their bodies (such as weight and body shape satisfaction, and sometimes referred to as the 'body concept') (Slade et al, 1990), and their beliefs about others' responses to their bodies (e.g. Hart et al, 1989). While these approaches differ in significant ways they share a concern with elaborating the relationship between the body and the self.

'Perceptual' body image research has explored people's accuracy in reporting their body size (e.g. Cappon and Banks, 1968), shape and weight in relation to their actual proportions (Cash et al, 1991; Slade, 1994); sensations or the lack of sensations from the body (Nijenhuis et al, 1996), as well as a person's general perception of their body boundaries (Murray and Gordon, 2001).

A second area of body image research concentrates upon person's evaluation of their bodies and its relationship to feelings of self-worth (Cash and Green, 1986). One early study in this regard is that of Secord and Jourard (1953) who introduced the term 'body cathexis' to separate people's subjective feelings about their bodies from their objective physical characteristics (Hart et al, 1989). This includes attitudes and beliefs about bodily appearance. This research commonly asks participants to evaluate their level of satisfaction with different bodily parts or properties (e.g. Secord and Jourard, 1953; Franzoi and Shields, 1984; Slade, Dewey, Newton, Brodie and Kiemle, 1990).

Another, less-well researched area of self-evaluated body experience is that of perceived physical competence, or people's feelings of confidence in demonstrating these skills in the company of others (Ryckman et al, 1982; Wylie, 1974). For example some people might be expected to feel more physically competent (such as athletes and people who engage in other physically demanding sports) while others might be expected to feel less competent (perhaps as the result of physical disability or chronic illness). As such, people's perceptions of their degree of physical competency might be expected to impact upon the relationship between their selves and their bodies.

The tendency by some people to perceive their bodies in an objectified manner has also been an area of research. McKinley and Hyde (1996) note how this tendency is most common in females, who have their embodiment partly shaped through a culture that objectifies the female physical form. The experience of the body as an object and the beliefs which support this experience are referred to by McKinley and Hyde (1996, p.183) as 'objectified body consciousness': "Women's relationship to their bodies becomes that of object and external onlooker: they exist as objects to themselves." Although McKinley and Hyde (1996) were particularly interested in female embodiment as a consequence of socialisation, we would theorise that objectified body consciousness would be a feature of people with a prior experience of an OBE (or who might experience one in the future).

The demonstrated relationship in the research literature between a person's negative evaluation of their body and low self-esteem (see Grogan, 1999 for a review) is the result of discomfort experienced because of a discrepancy between their actual (or perceived) bodily state and their preferred ideal state. We would characterise this phenomenon as one form of dissociation between the body and the self; one in which the body becomes objectified. Although research is lacking in the area of people's perceived physical competency and its relationship to the embodied self, we would hypothesise a similar dissociation between the self and body would occur for people who express low levels of such physical competency. It is such psychological distance between people's sense of self and their bodies that we hypothesise provides the psychological background to OBEs.

Increasingly, body image research is beginning to address the beliefs that people have about how other people will evaluate their bodies. This comes from recognition that satisfaction with one's appearance is highly influenced by social interaction. Hart et al. (1987), for instance, note that while many people think that others view their bodies favourably or are not concerned with others' responses to their bodies, others may be "chronically concerned" with how others evaluate their bodies. This may because their bodies are objectively unattractive or because they hold an unrealistically negative perception of their bodies. As a result they will tend to stay away from occasions and events (such as swimming in public) during which their bodies would be open to the (possible) evaluation of others.

Therefore, as with people's own affective responses, how they perceive others' perceptions of their bodies can be expected to have an impact upon their sense of embodied self. Such perceptions may be anticipatory, such as the person who avoids public displays of their physique due to an expected negative evaluation (whether expressed or not) by others. These concerns can be expected to lead to similar experiences to those resulting from a person's own negative evaluation of their bodies. That is, there is a discrepancy between the person's (perceived) bodily reality and a socially acceptable body. Again, we would argue that such discrepancy leads to a form of body-self dissociation, which in turn makes the occurrence of an OBE more likely.

Given that the phenomenological description of the OBE includes that of a dissociation, or separation, of the physical body and the self, we postulate that the person who experiences an OBE has a different relationship between their physical body and sense of self than do people without such experiences. We further postulate that the nature of this experience is a generalised dissociation (as compared with non-OBE experients) between their self and body that can be assessed on a number of levels: perceptual, affective and social. The present study, therefore, has a concern with the variety of body experiences that people have in relation to their reporting having had an out-of-body experience (we leave aside here any discussion of whether OBE experiences represent genuine OBE abilities). Several measures are employed in order to compare these experiences in an OBE and non-OBE sample. A number of hypotheses are made concerning the body experiences of those respondents reporting OBEs compared to those who do not:

1. They will experience higher levels of dissociation between their perceptual body and self.
2. They will have heightened self-awareness or self-consciousness.
3. They will have a reduced belief in their physical ability and in providing an acceptable physical self-presentation.
4. They will be more dissatisfied with their bodies.
5. They will have an objectified view of their bodies.
6. They will be more anxious at the prospect of having their physique evaluated by others.

METHODS

Participants

A total of 243 participants (171 females, 72 males) completed a web-survey regarding "body experience". Of these, 62, or 25% of respondents reported a previous OBE (45 females, 17 males, with a mean age of 29 years, SD = 10.9). In a review of the literature Alvarado (2000) reports an average OBE prevalence of 10% in the general population and 25% in student populations. Given that many of the respondents in the present study were or had been students the percentage of respondents in the present study reporting a previous OBE is comparable to Alvarado's provided average. All OBE respondents completed a question concerning how many OBEs they had had, although not everyone provided a precise figure. However, 79% of respondents reported having had two or more OBEs (52% reported 3 or more, 42% 4 or more, 39% 5 or more, 28% 6 or more, 22% 7 or more, 19% 8 or more, and 17% 10 or more). The non-OBE sample consisted of 181 participants (126 females, 55 males, with a mean age of 29 years, SD = 10.7).

Materials

Respondents completed a questionnaire comprised of 7 validated scales and one item (118 items in total) for assessing whether they had had a previous OBE. In the following we detail each of these questionnaire components in the order they were presented.

Measures

Somatoform Dissociation Questionnaire (SDQ-20): The SDQ-20 is a 20-item instrument designed to measure 'somatoform dissociation' or the degree to which the person experiences negative (e.g. losses of perceptions and control over functions) or positive (e.g. localized pain) perceptual or somatic symptoms indicative of dissociative disorder (Nijenhuis et al, 1996). Responses are made to 20 statements (e.g. "It sometimes happens that it is as if my body, or part of it, has disappeared") on a 5-point Likert scale (from 'not applicable' (1) to 'highly applicable' (5)) with the possible range of possible scores being 20-100. This scale is employed in the present study as an indication of respondents' perceptual body image. It was hypothesised that respondents with a prior OBE would score significantly higher on this scale than respondents without a prior OBE.

Self-Consciousness Scale (S-CS): The S-CS was constructed by Fenigstein, Scheier and Buss (1975) to assess individual differences in self-consciousness. In their construction of the scale the following behaviours were considered to constitute the area of self-consciousness: a preoccupation with past, present and future behaviour; sensitivity to inner feelings; recognition of one's positive and negative attributes; introspective behaviour; a tendency to picture one's physical appearance and presentation; and concern over the appraisal of others. The scale has three sub-scales: public self-consciousness (7- items), private self-consciousness (10-items) and social anxiety (6-items). The first of these refers to attending to one's inner thoughts and feelings (e.g. 'I reflect about myself a lot'), the second to one's awareness of their self as a social object (e.g. 'I'm self-conscious of the way I look'), and the third to discomfort experienced in the presence of others (e.g. 'Large groups make me nervous'). Therefore the full scale consists of 23 items (e.g. "I'm always trying to figure myself out") to which participants respond on a 5-point Likert scale ('extremely uncharacteristic' (0) to 'extremely characteristic' (4)). The range of possible scores for the complete scale and its sub-scales are (in order) 0-92, 0-28, 0-40, and 0-24. It was hypothesised that respondents with a prior OBE would score significantly higher on this scale and its subscales than respondents without a prior OBE.

Physical Self-Efficacy Scale (PS-ES): The PS-ES is a 22-item scale which assesses the degree to which respondents have a sense of physical self-efficacy (Ryckman, Robbins, Thornton and Cantrell, 1982). The scale has two subscales: Perceived Physical Ability (10-items, e.g. 'I am not agile and graceful') and Physical Self-Presentation (12-items, e.g. 'Sometimes I feel uncomfortable shaking hands because my hands are clammy'). High scores on the former are taken to indicate higher perceived physical ability, while high scores on the latter indicate greater confidence in the presentation of physical skills. Responses are made to 22 statements (e.g. "I find that I am not accident prone") on a 6-point Likert scale ('strongly agree' (1) to 'strongly disagree' (6)). The range of possible scores for the complete scale and its sub-scales are (in order) 22-132, 10-60, and 12-72. It was hypothesised that respondents with a prior OBE would score significantly lower on this scale and its subscales than respondents without a prior OBE.

Body Satisfaction Scale (BSS): The BSS was developed by Slade, Dewey, Newton, Brodie and Kiemle (1990) to measure satisfaction/dissatisfaction with body parts. The scale consists of 16 named body parts (e.g. "Legs") to which the respondent is asked to indicate on a 7-point Likert scale ('very satisfied' (1) to 'very unsatisfied' (7)) their degree of satisfaction with each. Three sub-scales are derived from this. The first is a 'general' scale which includes all items, the second a 'head' scale including most items relating to the face (7 items, e.g. 'jaw'), and the third a 'body' scale including those items relating to body parts (7 items, e.g. 'chest'). The scale was employed in the present study as a measure of respondents' affective responses to their own bodies. The range of possible scores for the complete scale and its sub-scales are (in order) 16-112, 7-49, 7-49. It was hypothesised that respondents with a prior OBE would score significantly higher on this scale and its subscales than respondents without a prior OBE.

Objectified Body Consciousness Scale (OBCS): The OBCS is a measure of objectified body consciousness developed by McKinley and Hyde (1996). The original scale is comprised of three sub-scales (8 items each): surveillance (viewing the body as an outside observer); body shame (feeling shame when the body does not conform); and appearance control beliefs. High scores on the Surveillance scale indicates the person watches their body frequently and thinks of their body in terms of how it looks rather than how it feels (e.g. 'I think more about how my body feels than how my body looks'). High scores on the Body Shame scale indicate that the person believes they are a 'bad person' if they do not fulfill cultural expectations for their body (e.g. 'I feel ashamed of myself when I haven't made the effort to look my best'). Finally, high scores on the Control Beliefs scale indicate that the person believes they are able to control their weight and appearance if they work hard enough (in contrast to low scorers who would believe that weight and appearance are controlled by factors such as heredity) (e.g. 'I really don't think I have much control over how my body looks'. The present study makes use of only the first two of these sub-scales. Participants responded to 16 statements (e.g. "I think more about how my body feels than how it looks") on a 5-point Likert scale ('strongly agree' (1) to 'strongly disagree' (5). The range of possible scores for the each of these sub-scales is 8-40. It was hypothesised that respondents with a prior OBE would score significantly higher on these subscales than respondents without a prior OBE.

Social Physique Anxiety Scale (SPAS): The SPAS is a 12-item self-report measure of the degree to which people become anxious at the prospect or presence of their physique being observed or evaluated by others (Hart, Leary and Rejeski, 1989). Respondents indicate the degree to which each statement (e.g. "It would make me uncomfortable to know others were evaluating my physique/figure") is characteristic of them on a 5-point Likert scale ('not at all' (1) to 'extremely characteristic' (5)). This scale was employed in the present study as it has a focus on how people feel others evaluate their personal appearance, and as such emphasises the social dimension of body image. The range of possible scores for the scale is 12-60. It was hypothesised that respondents with a prior OBE would score significantly higher on this scale than respondents without a prior OBE.

Belief in the Paranormal Scale (BPS): The BPS is an 8-item measure designed to assess respondents' level of belief in paranormal phenomena (Musch and Ehrenberg, 2002). Responses are made to eight statements (e.g. "I remember an event that I can only explain as a case of telepathy") on a 6-point Likert-scale ('total disagreement' (1) to 'total agreement' (6)). The range of possible scores on the scale is 8-48.

Item for Assessing the Occurrence of Out-Of-Body Experiences: In order to ascertain whether participants had experienced an out-of-body experience, respondents were provided with the following statement from Palmer (1979) and asked to indicate 'yes' or 'no': "Have you ever had an experience in which you felt that 'you' were 'outside of' or 'away from' your physical body; that is, the feeling that your consciousness, mind, or centre of awareness was at a different place than your physical body? (If in doubt, please answer 'no')."

Procedure

The survey was advertised using three methods; poster advertisements at Liverpool Hope University College (LHUC), e-mail advertisements to students and staff of LHUC, and a website advertisement on the LHUC virtual magazine (available in-house only). People were then free to respond in their own time by visiting the address of the web page provided in the advertisements.

The first page that they visited provided a brief outline of the study along with contact details for one of the authors (CM) and two questions concerning personal details (asking their gender and age). By using a button at the bottom of this page they were then taken to a new page presenting them with the first of the eight scales. Upon completion of each of the eight scales they moved to the next stage of the survey by using the button at the bottom of the page. Before proceeding to the next stage their submission was validated in order to ensure that valid responses had been provided (typically checking that the respondent had provided responses to all the items in the scale). If the validation failed they were presented with the page as they had submitted it, but with a message stating that there were errors in their submission, and the items requiring attention were highlighted. The software used for presenting these materials (written by JF) monitored the stage that participants had reached in the survey and prevented participants from returning to pages to make changes once they had provided a valid submission for that page.

Following the eight scales, the final page before the completion of the survey asked three questions relating to whether they had ever experienced being 'outside of' or 'away from' their body, and if so the frequency of these occurrences and whether they had felt either close to death or that they had died. They were also asked, whether they would like to be involved in future research, whether they would like to receive a copy of the findings, and, if appropriate, to provide contact information. Following submission of this form they were thanked for their participation.

RESULTS

Respondents' mean scores for each measure are shown in Table 1. Those respondents reporting a previous out-of-body experience (OBE) scored higher on The Somatoform Dissociation Questionnaire (SDQ-20) (M=35.14, compared to 26.15), the Self-Consciousness Scale (S-CS) (M=55.04, compared to 50.69), the Social Physique Anxiety Scale (SPAS) (M=37.32, compared to 35.98), and on the Belief in the Paranormal Scale (BPS) (M=33.77, compared to 24.8) than those not reporting a previous OBE. They also scored lower on the Physical Self-Efficacy Scale (PS-ES) (M=78.02, compared to 81.80). Scores on the two subscales of the Objectified Body Consciousness Scale (OBCS) were similar (for Surveillance the OBE sample had mean of 22.29, compared to 22.68, and for Body Shame they had a mean of 27.56 compared to 27.5).

Table 1: Mean Scores (and Standard Deviations) by Respondents on Questionnaire Scales

Measure	Non-OBE Group (n=181)	OBE Group (n=62)
The Somatoform Dissociation Questionnaire (SDQ-20)	26.15 (7.24)	35.14 (14.43)
Body Satisfaction Scale	47.73 (15.74)	56.53 (22.44)
Head	19.34 (7.69)	23.74 (10.51)
Body	23.44 (8.52)	26.53 (10.61)
Self-Consciousness Scale (S-CS)	50.70 (13.91)	55.04 (14.16)
Private Self-consciousness	21.69 (6.71)	24.71 (6.99)
Public Self-consciousness	16.85 (6.21)	17.60 (6.28)
Social Anxiety	12.16 (5.70)	12.74 (6.06)
Physical Self-Efficacy Scale (PS-ES)	81.80 (14.70)	78.02 (13.51)
Perceived Physical Ability	35.14 (9.33)	34.37 (6.78)
Physical Self-Presentation	46.66 (8.53)	43.65 (9.40)
Objectified Body Consciousness Scale (OBCS)		
Surveillance	22.68 (6.49)	22.29 (7.11)
Body Shame	27.50 (7.14)	27.56 (7.26)
Social Physique Anxiety Scale (SPAS)	35.98 (8.69)	37.32 (9.83)
Belief in the Paranormal Scale (BPS)	24.40 (8.44)	33.77 (8.47)

The results of ANOVA significance tests are shown in Table 2. Respondents reporting a previous OBE were found to score significantly higher on The Somatoform Dissociation Questionnaire ($F(1, 241) = 40.61$, $p<.001$), the Self-Consciousness Scale ($F(1, 241) = 4.48$, $p=.035$) and its Private Self-consciousness subscale ($F(1, 241) = 9.18$, $p=.003$), the Body Satisfaction Scale ($F(1, 241) = 11.45$, $p=.001$) and both its Head ($F(1, 241) = 12.43$, $p=.001$) and Body ($F(1, 241) = 5.3$, $p=.022$) subscales, and the Belief in the Paranormal Scale

(F(1, 241) = 56.71, p<.001) than respondents without a previous OBE. OBE respondents also scored significantly lower on the Physical Self-presentation subscale of the Physical Self-efficacy Scale (F(1, 241) = 5.46, p=.02). There were no significant differences between these groups on the complete Physical Self-Efficacy Scale or its Perceived Physical Ability subscale, the Surveillance or Body Shame subscales of the Objectified Body Consciousness Scale, or the Social Physique Anxiety Scale.

Table 2: Results of ANOVA Significance Tests on Experimental Measures

Measure		Sum of	df	Mean	F	Sig.
Somatoform Dissociation	Between	3732.745	1	3732.745	40.61	<.001
	Within Groups	22153.362	241	91.923		
	Total	25886.107	242			
Self-Consciousness Scale	Between	874.767	1	874.767	4.48	.035
	Within Groups	47057.142	241	195.258		
	Total	47931.909	242			
Private Self-consciousness	Between	422.472	1	422.472	9.18	.003
	Within Groups	11085.824	241	45.999		
	Total	11508.296	242			
Public Self-consciousness	Between	25.697	1	25.697	.66	.416
	Within Groups	9339.892	241	38.755		
	Total	9365.588	242			
Social Anxiety	Between	15.627	1	15.627	.46	.496
	Within Groups	8096.225	241	33.594		
	Total	8111.852	242			
Physical Self-Efficacy Scale	Between	659.662	1	659.662	3.18	.076
	Within Groups	50060.420	241	207.720		
	Total	50720.082	242			
Perceived Physical Ability	Between	27.179	1	27.179	.35	.552
	Within Groups	18470.015	240	76.639		
	Total	18497.193	241			
Physical Self-Presentation	Between	419.044	1	419.044	5.46	.020
	Within Groups	18496.956	240	76.639		
	Total	18916.000	241			
Body Satisfaction Scale	Between	3578.682	1	3578.682	11.45	.001
	Within Groups	75305.170	241	312.470		
	Total	78883.852	242			
Head	Between	896.066	1	896.066	12.43	.001
	Within Groups	17366.313	241	72.059		
	Total	18262.379	242			
Body	Between	437.997	1	437.997	5.30	.022
	Within Groups	24781.769	241	102.829		
	Total	25419.193	242			
Objectified Body-Consciousness Scale	Between	6.997	1	6.997	.16	.691
Surveillance	Within Groups	10648.189	241	44.183		
	Total	10655.185	242			
Body Shame	Between	.176	1	.176	.00	.953
	Within Groups	12396.491	241	51.438		
	Total	12396.667	242			
Social Physical Anxiety Scale	Between	83.503	1	83.503	1.03	.311
	Within Groups	19509.460	241	80.952		
	Total	19592.963	242			
Belief in the Paranormal Scale	Between	4050.535	1	4050.535	56.71	<.001
	Within Groups	17214.585	241	71.430		
	Total	21265.119	242			

Discussion

As predicted, respondents reporting a previous OBE were found to experience higher levels of dissociation between their perceptual body and self, to have a heightened self-awareness or self-consciousness, to be more dissatisfied with their bodies, and to have lower confidence in the presentation of their physical skills. However, the hypotheses that they would have a reduced belief in their physical ability, an objectified view of their bodies, and be more anxious at the prospect of having their physique evaluated by others were not supported.

The first of these findings is in agreement with Irwin's (2000) study, in which scores of somatoform dissociation (using the same scale employed in the present study) were found to independently discriminate between people with and without a prior OBE, as well as the only independent variable which contributed significantly in predicting OBE frequency. Irwin's own rationale for studying somatoform dissociation in OBEs was due to the apparent phenomenological separation of perceptual experience and the person's sense of self or identity which occurs during an OBE. As discussed earlier, on one level this would also appear to support the psychological theories of OBEs advanced by Palmer (1978) and Blackmore (1984) in which the loss of somatosensory information from the body precipitates an OBE. However, the scale used in the present study also included items which referred to amplified somatosensory experiences which are more difficult to explain by reference to these theories alone[1]. (One issue worthy of note is that the mean scores by respondents in Irwin's study and the present study differ; the mean score of 26.15 by the non-OBE group in the present study is larger than that for both the OBE group (25.91, SD 8.13) and the non OBE group (21.74, SD 2.59) in that of Irwin's).

The heightened self-consciousness found in the OBE sample was expected due to the theorised dissociation between the self and body that this sample was expected to have; that is, the self was expected to increase in thematic awareness as the person experienced a greater degree of dissociation between their body and self. This is in contrast to persons who did not experience a dissociation between their self and body, but as suggested in the literature on the phenomenology of embodiment, were expected to experience their bodies largely in a pre-reflective manner with their 'self' oriented (through the medium of the body) 'outwards'. In their development of the Self-consciousness Scale, Fenigstein et al. (1975) suggested people who scored high on this measure would be better able to take part in meditation. Whereas people without this psychological and phenomenological separation of self and body were expected to have a stronger sense of embodiment, the OBE sample was expected to have a weak sense of embodiment in part characterised by an increased focus on the private or inner self. Indeed, when analyses were carried out on the sub-scales of the Self-Consciousness Scale, namely the Private Self-Consciousness, Public Self-Consciousness and Social Anxiety sub-scales, only the first of these remained statistically significant, that is the OBE group was higher in Private Self-Consciousness (although the mean for the OBE group was higher on all three measures). That no differences were found on the Public Self-Consciousness sub-scale is surprising given that OBE participants scored significantly lower than non-OBE participants did on their level of confidence in the presentation of physical skills. However, it should be borne in mind that the reduced number of items that each of the sub-scales has, compared to the full scale, may in part explain this.

The higher level of body dissatisfaction in the OBE sample had been expected and was a feature which we theorised characterised a general dissociation between body and self by OBE experients; that is our reading of the literature on the phenomenology of embodiment had lead us to expect that people who were dissatisfied with their bodies would not have the close identification between body and self found in persons with high levels of body satisfaction. However, we had also expected that lower levels of physical self-efficacy

[1] One of the reviewers suggested that an analysis should be performed on those items on the SDQ that referred to these amplified or 'positive' symptoms to confirm our contention that they contribute significantly to a relationship between somatoform dissociation and out-of-body experiences. We identified five of the 20 items which qualify in this regard (items 5, 12, 14, 15 and 17) (some other items refer to both or are difficult to place in either 'deficit' or 'positive' categories). The mean score for OBE experients on all these items was higher, and a oneway ANOVA for these items confirmed that the OBE group scored significantly higher on these items than the non-OBE group ($F(1, 241) = 27.225$, $p>.001$).

would characterise people with a weak sense of embodiment, and hence people with previous OBEs, as well as expecting higher levels of social physique anxiety. However, there was no support for any differences between these groups on either of these measures.

The finding that there are no differences between the groups concerning having an objectified view of their bodies or at the prospect of having their physique evaluated by others are surprising, particularly given the differences between the groups on levels of body dissatisfaction. However, as stated earlier, there were also no differences between the groups on the Public Self-consciousness and Social Anxiety sub-scales on the Self-consciousness Scale, suggesting that as well as having no differences in social anxiety in relation to their physique, these groups did not differ in regards to public self-consciousness and social anxiety in general.

We believe the present findings are informative in understanding the occurrence of OBEs. They point to a relationship between a broader range of body experiences (and different conceptualizations of 'body image') and OBEs than investigated in extant literature. We feel that further study of the different facets of body image as outlined in this paper may prove informative about such issues as the prevalence and frequency of OBEs, as well OBE features. However, further work must also address the limitations of the present study; in particular such work needs to be carried out with other populations (in contrast to our focus on an academic population), and the various forms of body image related in this paper should be examined in relation to different forms or types of OBEs (rather than the broad delineation in the present study between those who responded 'yes' or 'no' to the item for assessing whether an OBE had occurred).

To conclude, the findings within this paper lend support to our contention there are pre-existing differences in the body experiences of those people who do and do not have OBEs which are not accounted for in current psychological theories. (In the present study, as in previous work, OBE respondents tended to have had multiple OBEs with only 21% reporting having had only one). We would argue that it is the exacerbation of these pre-existing differences which results in an OBE, although further research is required to support this contention.

ACKNOWLEDGEMENTS

We would like to thank two anonymous reviewers for their helpful and supportive critiques of an earlier version of this paper. We are also very grateful to the Bial Foundation who provided financial support for this study as part of the project 'The Flexibility of Body Boundaries and its Relationship to Out-of-Body Experiences' (number 124/02).

References

Alvarado, C.S. (1986) Research on Spontaneous Out-of-body Experiences: A Review of Modern Developments, 1960-1984. In Shapin, B. and Coly, L. (Eds.) *Current Trends in Psi Research.* New York: Parapsychology Foundation. (pp.140-167).

Alvarado, C.S. (1992) The Psychological Approach to Out-of-body Experiences: A Review of Early and Modern Developments. *Journal of Psychology, 126(3),* 237-240.

Alvarado, C.S. (2000). Out-of-body experiences. In Cardeña, E., Lynn, S.J. and Krippner, S. (Eds.) *Varieties of Anomalous Experiences.* Washington, DC: American Psychological Association. (pp. 183-218).

Blackmore, S.J. (1984) A Psychological Theory of the Out-of-body Experience. *Journal of Parapsychology, 48,* 201-218.

Cappon, D. and Banks, R. (1968) Distorted Body Perception in Obesity. *Journal of Nervous and Mental Disorders, 46,* 465-467.

Cash, T.F., Wood, K.C., Phelps, K.D. and Boyd, K. (1991) New Assessment of Weight-Related Body Image Derived From Extant Instruments. *Perceptual and Motor Skills, 73,* 235-241.

Cash, T.F. and Green, G.K. (1983) Body weight and body image among college women: Perception, cognition, and effect. *Journal of Personality Assessment, 50,* 290-301.

Fenigstein, A., Scheier, M.F. and Buss, A.H. (1975) Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43(4),* 522-527.

Franzoi, S. and Shields, S.A. (1984) The Body Esteem Scale: Multidimensional Structure and Sex Differences in a College Population. *Journal of Personality Assessment, 48(2),* 173-178.

Grogan, S. (1999) *Body Image: Understanding Body Dissatisfaction in Men, Women and Children.* London: Routledge.

Hart, E.A., Leary, M.R., Rejeski, W.J. (1989) The Measurement of Social Physique Anxiety. *Journal of Sport and exercise Psychology, 11,* 94-104.

Irwin, H.J. (1985) *Flight of Mind: A Psychological Study of the Out-of-Body Experience*. Metuchen, NJ: Scarecrow Press.

Irwin, H.J. (2000) The Disembodied Self: An Empirical Study of Dissociation and the Out-Of-Body Experience. *Journal of Parapsychology, 64(3),* 261-276.

Secord, P.F. and Jourard, S.M. (1953) The Appraisal of Body-Cathesis: Body-Cathesis and the Self. *Journal of Consulting Psychology, 17,* 343-347.

McKinley, N.M. and Hyde, J.S. (1996) The Objectified Body Consciousness Scale. *Psychology of Women Quarterly, 20,* 181-215.

Murray, C.D. and Gordon, M. (2001) Changes in Bodily Awareness Induced by Immersive Virtual Reality. *CyberPsychology and Behavior, 4(3),* 365-372

Musch, J. and Ehrenberg, K. (2002) Probability Misjudgement, Cognitive Ability, and Belief in the Paranormal. *British Journal of Psychology, 93,* 169-177.

Nijenhuis, E.R.S, Spinhoven, P., van Dyke, R., van der Hart, O. and Vanderlinden, J. (1996) The Development and Psychometric Characteristics of the Somatoform Dissociation Questionnaire. *The Journal of Nervous and Mental Disease, 184(11),* 688-694)

Palmer, J. (1978) The Out-of-body Experience: A Psychological Theory. *Parapsychology Review, 9(5),* 19-22.

Palmer, J. (1979) A Community Mail Survey of Psychic Experiences. *Journal of the American Society for Psychical Research, 73,* 221-251.

Ryckman, R.M., Robbins, M.A., Thornton, B. and Cantrell, P. (1982) Development and Validation of a Physical Self-Efficacy Scale. *Journal of Personality and Social Psychology, 42,* 891-900.

Slade, P. D., Dewey, M. E., Newton, T., Brodie, D., and Kiemle, G. (1990) Development of the Body Satisfaction Scale (BSS). *Psychology and Health, 4,* 213-226.

Address for correspondence: Craig D. Murray,
Psychology Department, Liverpool Hope University College, Hope Park, Liverpool L16 9JD, UK.
E-mail: murrayc@hope.ac.uk

ON THE FALSE HYPOTHESIS OF PSI-MEDIATED SHIFT OF STATISTICAL AVERAGE IN TESTS WITH RANDOM EVENT GENERATORS

Fotini Pallikari
Faculty of Physics, University of Athens

ABSTRACT

The micro-psychokinesis (micro-PK) hypothesis states that intention can mentally shift the statistical average of the outcomes of an inherently random process in the desired, pre-stated direction. The most recent meta-analysis based on all available to date micro-PK data, testing the direct influence of human intention on the outcomes of true random number generators (RNG), does not support this hypothesis. Furthermore, a very large-scale experiment set out by a consortium of independent research groups to replicate the micro-PK hypothesis, has failed to show the effect. Either micro-PK is not a real effect, or the micro-PK hypothesis has been wrongly considered at its naissance to introduce mean-shifts. It is therefore proposed that the definition of micro-PK will have to change. In this work the implications of the current evidence will be shown to support the second alternative, as long as micro-PK prolongs the length of runs of successful and failed trials indistinguishably, a case of persistence of like binary events. The result of such persistence in the length of a run will appear as a broadening of the funnel plot, the graph which represents the scatter of size of study, N, against the proportion of successes in it, pi. The mechanism that is compatible with the current evidence around RNG-micro-PK is shown that cannot constitute a real separate process mixed with the random process which either adds a fixed proportion of successes to it, or shifts the statistical average by a fixed number of standard errors. This paper puts forward the hypothesis that, provided there is a micro-PK effect to explain, it should be sought after in the broadening of variance and not in the shift of mean. The proposed mechanism, which has been given the name 'psi-gluing process' and which prolongs the length of the randomly generated runs of successes and failures, can adequately account for the shape of the micro-PK funnel plot, its convergence to 0.5 and its broadened width. It can also account for other relevant evidence such as the decline effect and the rendering of the statistical distribution of all large number of data in a micro-PK database to the perfect Gaussian. In the long run the statistical average of data in a micro-PK-RNG test will have to converge to chance. This hypothesis is accompanied by a prediction concerning the broadened boundaries that border the scatter of micro-PK scores per study size, a prediction which further experiments can either confirm or falsify. It is proposed here that the psi-gluing process is fundamental to the mind-matter coupling mechanism and therefore present in other related phenomena such as remote viewing or telepathy.

INTRODUCTION

Micro-PK has been defined as the potential of human intention to shift the statistical average of data generated by a random process in the desired, and pre-stated direction either above or below the theoretically expected value (Glossary, 2001). Unfortunately, the database of numerous micro-PK experimental results accumulated over the last sixty years does not support this hypothesis (Steinkamp, Boller & Bösch, 2002) and as individual example one can cite the outcome of a micro-PK test with Schmidt machines, by an experimenter other than Schmidt (Pallikari, 1997). Furthermore, a carefully designed micro-PK experiment performed by a consortium of independent research teams yielded no evidence in support of the micro-PK hypothesis (Jahn, Mischo, Vaitl et al, 2000). If intention could shift the mean of random processes, then all experiments measuring physical properties should be subjected to it. Repeatable independent measurements of physical properties performed at high precision in prestigious laboratories would have a statistical average which shifts from the theoretically expected according to intention. Spectroscopy experiments, for instance, where billions of photons are involved, would display frequency shifts by mental intervention alone. Not only this has not been observed over experiments taking place all the time in research laboratories, to the best of my knowledge, but also a micro-PK experiment involving photons has failed to support the micro-PK hypothesis (Jeffers & Sloan, 1992) even though the same experiment and the same apparatus at the PEAR lab had recorded a significant score at the 5% level with a terminal z-score of 1.654 (Ibison & Jeffers, 1998). Clearly, the current evidence flies in the face of previous claims (Radin & Nelson, 2002) regarding mentally induced micro-PK mean shifts. The information provided by the most recent meta-analysis (Steinkamp, Boller & Bösch, 2002) looks at the RNG-micro-PK database from a different and more interesting angle, one that meta-analyses carried out earlier by other, yet related, research teams have not considered. This most recent meta-analysis indicates a great deal weaker micro-PK effect than what the previous analyses have estimated. It shows that the estimated weak strength of effect can be rendered to non-existence easily, by mere less than seventy unpublished studies.

How can the paradox around the micro-PK hypothesis be resolved? We suggest two remedies: either by abandoning the idea that micro-PK is a real effect, or by devising a viable explanation in its support, one that agrees with the current evidence around it. In fact, in this paper we shall present a hypothesis that keeps the door still open to the likelihood of its existence. This paper claims that there is a micro-PK effect to explain, that it is found in the funnel plot statistical heterogeneity, the broadening of variance, and not in a shift of mean according to the held the belief so far. It does not claim that micro-PK is just data selection and filedrawer effect and also provides a prediction regarding the boundaries in which the proportion of successes in a micro-PK test range per study size. In the following section we shall review the current evidence around micro-PK and its implications, while attempting to provide a viable explanation for it.

THE FUNNEL PLOT IN THE MICRO-PK META-ANALYSIS

As was mentioned in the introduction, a carefully designed and thorough meta-analysis was published two years ago to test the hypothesis of possible correlations between direct human intention and the concurrent output of true RNGs (Steinkamp, Boller & Bösch, 2002). The analysis was performed at the level of bits by appropriately converting the result of a study, wherever that was allowed, into a percentage of bits generated overall in the direction of intention. Thus, the effect size, symbolized in the article by 'pi', was a parameter measuring the proportion of successful trials in a study and varying between 0 and 1. Values of pi above 0.5 indicated that the study's score was in the direction of the pre-stated intention to influence the statistics of the random process since the proportion of successful trials was over 50%. Any score below 0.5 indicated that the result was against the direction of intention. The size of study (in logarithmic scale) was then plotted against pi and the graph took the shape of an asymmetrical inverted funnel, figure 1. Such graphs, called funnel plots, are popular mainly in medical sciences and in many occasions can be useful to identify the presence of bias in the database and help evaluate the validity of the meta-analysis. The funnel plot owes its shape to the fact that as the size of a study increases so does the precision in estimating the underlying effect under investigation. Results from small studies will scatter largely at the bottom of the graph, while the spreading will narrow at the top for larger studies to converge at the value which truly represents that underlying effect. This will make the plot of study size against effect size to look like an inverted funnel. The funnel plot of an unbiased and adequately large database is expected, for that reason, to have a symmetrical shape due to the natural scatter of measurements of an effect about its real value. The individual scores will be normally distributed about the true effect size and the 95% of them are expected to scatter within 1.96 standard errors at either side, equation 1 and A-4. On the contrary, funnel plots, which are asymmetrical and skewed, indicate the presence of bias, implying that studies were not included in the plot for a number of possible reasons. A rudimentary, simple approach to observe the funnel plot asymmetry is visually.

The funnel plot of the most recent meta-analysis (Steinkamp, Boller & Bösch, 2002) indicates that (a) the underlying micro-PK RNG effect converges to 0.5, a value expected by chance alone and against the micro-PK hypothesis, and (b) the database is highly asymmetric and therefore biased. In the next sections, we shall expand further on these two features. Other members in the parapsychology community discuss size dependent effect size models other than the obvious dependence of pi on N in equation 1. Yet, that is not the point of this paper, as the approach followed here sees no indication for it in the funnel plot.

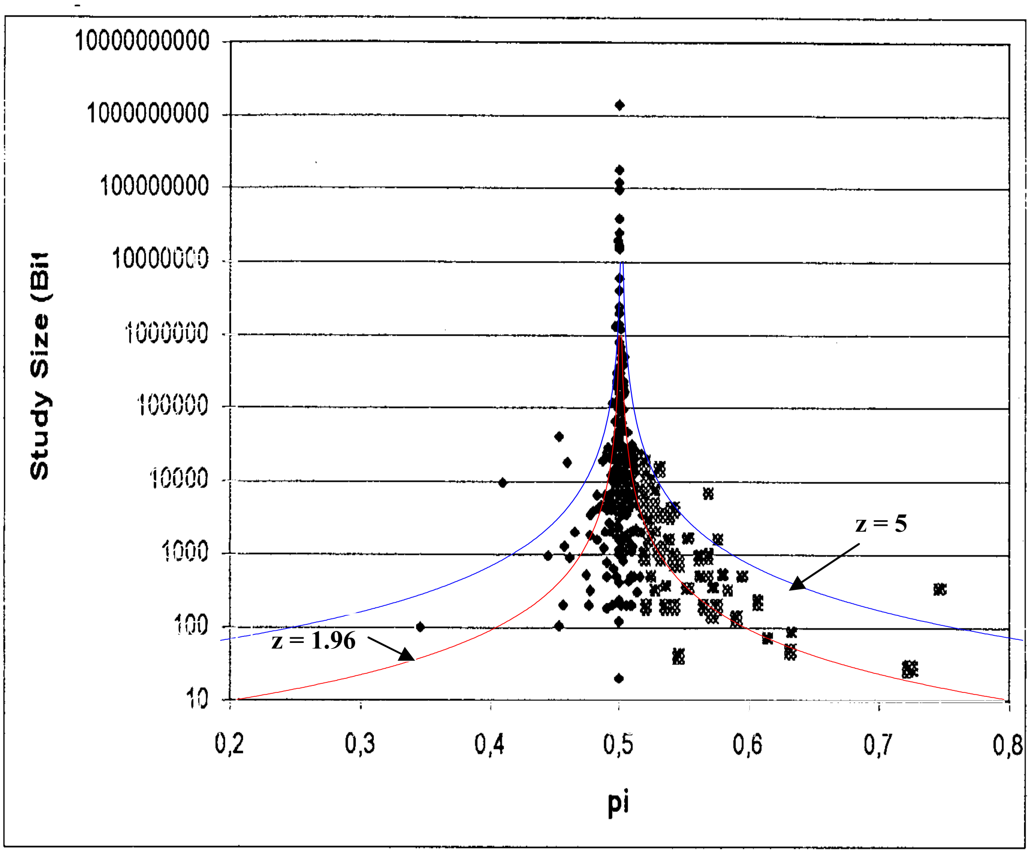


Fig. 1: The funnel plot of the meta-analysis of PK-RNG data (Steinkamp, Boller & Bösch, 2002)[1]. The curve obeying equation 1, for z = 1.96, only envelopes about 50% of all data. A wider ranging and more enclosing envelope curve requires z ˜ 5.

EXPLORING DIFFERENT INTERPRETATIONS FOR THE FUNNEL PLOT OF THE MICRO-PK META-ANALYSIS

Consider a set of random unbiased proportions[2], pi, of successes and failures to display an effect in a Bernoulli process[3] (see also appendix). The spread of 95% of them, centered at 0.5 and having standard error $\frac{0.5}{\sqrt{N}}$, should be found lying under the envelope curve determined by the general formulas of equation 1 below for z=1.96 also plotted in figure 1

[1] The pi coordinate of a few data located at extreme positions on the left half of the original funnel plot was later corrected by one of the authors (Bösch).

[2] Assuming that the proportions of successes, constituting a large database of independent studies, is normally distributed (Spiegel, 1961).

[3] Testing the probability that a particular event will happen in any single trial.

$$\boxed{pi = 0.5 \pm z\frac{0.5}{\sqrt{N}}} \qquad \text{or} \qquad \boxed{N = \frac{(0.5z)^2}{(pi-0.5)^2}} \tag{1}$$

To test the validity of equation 1, one million random bits were generated in the computer. They all fitted well under the z =1.96 envelope curve and were symmetrically spread about the value 0.5, thus confirming this theoretical prerequisite. But not only the computer-simulated data, but also the control data of the same meta-analysis confirmed equation 1 for z=1.96 (Bösch, 2003), even if they were fewer in number than the experimental ones.

However, the experimental data of the recent meta-analysis are asymmetrically spread and scattering strongly about 0.5, figure 1. Roughly, only half of them are enveloped under the expected 95% confidence level (z = 1.96), a behaviour both the computer-simulated and meta-analysis control data exhibit. Of the set of curves obeying equation 1, the one that would more adequately envelope them, would be broader requiring a z-value equal to approximately 5, figure 1. Having drawn this broader envelope curve, we observe that the distribution of experimental data under it is visibly denser in the right half of the funnel plot for pi above 0.5. The section of the funnel plot on the left half between coordinates $10<N<100$ and $0.38<pi<0.47$ or between coordinates $10^3<N<10^{4.5}$ and $0.47<pi<0.5$ clearly displays areas void of data points. We shall attempt different interpretations for this funnel plot asymmetry.

The funnel plot is not the result of a separate random process overlapping with a biased micro-PK process.

Suppose the experimental data were the product of more than one underlying processes, one of which is the main process generating random events and the other the micro-PK process which shifts the average from that expected by chance, 0.5, but still generates normally distributed events. Just by chance micro-PK data will be generated scoring a value centered at 0.5 and spread about this mean (by 95%) as far to the right as to the left by $1.96\frac{0.5}{\sqrt{N}}$ according to equation 1. The micro-PK biased data are spread about a pi>0.5 (to agree with the micro-PK hypothesis) and the 95% of them would lie under an envelope curve of width again $1.96\frac{0.5}{\sqrt{N}}$ at each study size N, since the micro-PK data are still normally distributed. Additionally, we assume that the random and micro-PK modulated data distributions overlap to account for the asymmetry of the funnel plot of figure 1. The evidence cannot substantiate such postulation.

Firstly, such assumption implies that there must be not just one but many underlying overlapping biasing processes of different strengths, each shifting the average to the right of 0.5, to account for the asymmetric spread of experimental data in the funnel plot. Most importantly, each of these underlying micro-PK effects must become more prominent on the plot, acquiring greater accuracy, as the study size becomes larger. In other words each micro-PK process must be identified on the funnel plot by a separate narrow peak, as figure 2 illustrates. Otherwise, one should conclude that the underlying micro-PK mechanism weakens as the sample size increases above about N=1,000 and completely disappears at N=10,000 or so. Such assumption implies that whatever individual pi-shifting effect creates the high density of data, on the right half of the funnel plot, it cannot represent a real mechanism as that cannot withstand the diluting effect of the law of large numbers that renders it to chance.

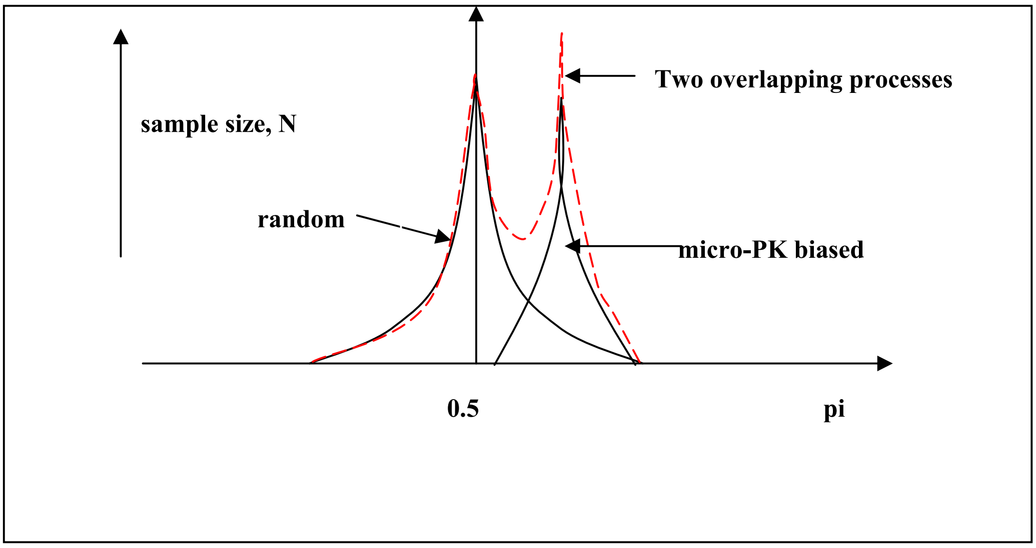


Fig. 2. Simulation of the funnel plot drawn as sample size against proportion of successes, pi, in a test biased by micro-PK. Two processes are present, the random (centered at 0.5) and the biased micro-PK (shifted to the right). Their overlap is shown in dashed line displaying two clear peaks. If a micro-PK process exists shifting the mean above pi= 0.5, it should become more prominent as it acquires higher precision at larger sample sizes.

Finally, the presence of experimental data beyond the z=1.96 envelope to the left half of the funnel plot demands that, once such a micro-PK process has been considered viable for the right half, it should also operate at directions against the pre-stated intention and therefore against the micro-PK hypothesis. Each of these mechanisms should also become distinguishable as narrow peaks in the funnel plot, but now shifted to effect sizes pi<0.5. The shape of the funnel plot obviously does not support such a supposition, as there is only one peak present and there are void areas on the left below 0.5. We conclude that we cannot assume the existence of different overlapping pi-shifting processes with the random process to account for the shape of the meta-analysis funnel plot of figure 1.

The micro-PK effect does not shift the statistical average of the random process by a fixed z value.

As the assumption of mean shifted overlapping funnel plots cannot be sustained, we consider alternative explanations to explain the high density of experimental data extending beyond the z=1.96 envelope. We suggest this time that the micro-PK data may have broadened the funnel plot by shifting their randomly generated outcomes by several standard errors, five for instance. In other words, the average proportion of successes is 0.5, as predicted by chance, while the individual micro-PK outcomes range between the limits set by equation 1 for z=5. As the mean shift operates now in both directions, above and below 0.5, the funnel plot becomes wider. Can such broadening mechanism be backed by evidence?

Can, for instance, the micro-PK mechanism be accredited to shift the average proportion of successes in a set of studies of size N, from the expected $\overline{pi} = 0.5$, to e.g. $\overline{pi} = 0.5 + \frac{0.5 \cdot 5}{\sqrt{N}}$, that is shift the average by a fixed number, z=5, of standard errors? This would imply, for example, that for a study size N=10^9 trials, the average percentage of successes in the direction of intention is 0.50008, something that could be supported by the shape of the funnel plot. It will also imply that the average pi for N=10^3 is 0.579, but this is clearly not supported by the shape of the funnel plot. If the funnel plot converges to 0.50008, provided that the database is normally distributed as expected for N>30 (Spiegel, 1961), the 95% of studies included in the funnel plot will have pi values spread in the range $0.50008-1.96\frac{0.5}{\sqrt{N}} < pi < 0.50008+1.96\frac{0.5}{\sqrt{N}}$. But the actual funnel plot data do not spread about 0.50008 by $\left(1.96\frac{0.5}{\sqrt{100}} = \right)0.098$ at N=100, for instance. Then again, the average pi of studies having size N=1,000 is clearly not 0.597. We cannot assume, therefore, that there is a PK mechanism that shifts the statistical average of a random process after N events were generated at a constant z value, e.g. z =5, since this cannot be supported by the evidence that the meta-analysis provides. What is obvious from the graph is that only the scatter of individual test outcomes has been broadened according to equation 1, where z=5.

The broadening of funnel plot is not due to the addition of a fixed proportion of micro-PK successes.

If the micro-PK mechanism does not shift the average pi in N trials by a fixed number of standard errors, z, could we assume that it adds to the expected average a fixed proportion of micro-PK biased successes instead? Will a certain proportion of successes in a study of N trials, be entirely introduced by the micro-PK biasing mechanism, while the rest be due to the random process?

In the previous section we have seen that the shift to 0.50008 (from 0.5) at 10^9 trials, of high statistical significance, could be acceptable on the basis of the funnel plot shape. Yet, the actual data spread about it would have to be less wide than actually is. Is the score average at every study size N shifted by 0.00008? The average pi must then be the sum pi_R of randomly generated successes and pi_{PK} of micro-PK generated successes

$$\overline{pi} = \overline{pi}_R + \overline{pi}_{PK} \tag{2}$$

The funnel plot clearly does not support this assumption as the density of data points at N smaller than 10^4 is centered at pi $\neq$0.50008 and closer to 0.52. We cannot therefore assume that the micro-PK mechanism adds into a sequence of successes and failures a fixed percentage of successes on top of those generated by chance. The funnel plot does not support such a hypothesis.

WHAT CAN CAUSE THE PRESENT ASYMMETRY IN THE MICRO-PK-RNG FUNNEL PLOT?

Since the meta-analysis database was put together under strict conditions to examine the correlation between direct human intention and the concurrent output of true RNGs, excluding telepathy studies, animals or plants, ambiguously intentional effects, or retro-PK, it can be considered sufficiently homogeneous. The Steinkamp et al paper makes this clear on page 262:

> After this detailed search, the meta-analysis presented here focused specifically on experiments examining the correlation between *human intention* and the *concurrent* output of true RNGs, in the hope of obtaining a more homogeneous data set.

Yet, in the meta-analysis the data can be considered heterogeneous in another sense, that of statistical heterogeneity. This type of heterogeneity is tested with the appropriate statistics related to chi-square tests (Deeks, 2001 & 2003). Statistical heterogeneity manifests itself whenever the pi scores are more different from each other than one would expect by chance. This will make the scatter of data on the funnel plot broader than expected by chance and is obviously related to the data variance. The experimental data in the Steinkamp meta-analysis were identified to be statistically heterogeneous by the chi-square measure, thus quantifying the otherwise apparent broadening of the funnel plot, compared to what chance would predict for the 95% of data (figure 1). When explanations for the funnel plot broadening cannot be found, it is customary to consider it caused by a normal distribution of effect sizes, if that can be justified. This approach, which acknowledges the presence of heterogeneity, is known as the 'random effects' estimate of statistical heterogeneity (Cooper, 1994). Alternatively, one could assume that the true effect size is represented by just one value in every study, i.e. it is fixed across studies. This alternative approach constitutes the 'fixed effect' approach to statistical heterogeneity, which actually assumes that the observed differences in the scatter of data are somehow attributed to chance.

Except for the most recent meta-analysis, other efforts to correct for the missing data have been implemented in the past by related research teams introducing the file drawer effect correction to conclude that the number of missing studies were impossibly high (Radin & Nelson, 2002). If a funnel plot had been drawn it would have been found extremely asymmetric. But, scientists do not agree with either the efficiency of the filedrawer approach (Scargle, 2000) or the magnitude of its significance (Steinkamp, Boller & Bösch, 2002). In any case, it would be wiser not to use a controversial statistical approach, such as the file drawer correction, in support of a controversial phenomenon, such as the influence of human intention on the statistics of a random process.

It should be noted that the Steinkamp et al paper suggests a list of causes of statistical heterogeneity, such as selection biases, true heterogeneity, data irregularities and chance. It was shown in a work within biology meta-analyses that the publication bias, as one of the possible selection biases, could substantially distort the estimated effect size without that to harm the superiority of the meta-analysis as a reviewing technique (Jennions, 2004). Steinkamp et al also showed that the superiority of selected star participants, one of the true heterogeneity factors, couldn't be the cause of funnel plot asymmetry. What about the question of data irregularities as the source of bias? Regarding the poor methodological design factor, it was shown that small studies report observations against the hypothesis less often than would have been expected in normal variation; more often small studies are reported when they present an effect in favour of the hypothesis (Palmer, 2002) and that introduces a publication bias and a funnel plot asymmetry. The funnel plot of Steinkamp et al shows clear signs of such areas void of data in the small studies region, small N. All the above thoughts lead one to consider, if not the only one, the major factor for funnel plot asymmetry is the selection bias caused by unreported studies. If these studies could be identified they would be expected to fill the present voids data in the funnel plot between N=100 and 10,000 since some data already exist in the left hand side and beyond the z=1.96 envelope curves. Therefore, the presence of unreported data points would not weaken the statistical heterogeneity, the broadening of the funnel plot, the variance, but enhance it.

Once the publication bias is accounted for as the main cause of the funnel plot asymmetry there remains to speculate on the possible broadening mechanism. Certain forms of fixed effect mechanisms have been accounted for in the previous sections and were rejected as unsuitable. This paper acknowledges the presence of statistical heterogeneity and suggests that the funnel plot broadening mechanism is the psi-induced persistence of successes or failures (psi-gluing). It assumes that its effect will approximately be represented by a normal distribution of pi values that do not exceed the N=f(pi) curve determined by about z=5.4 at every study size N. As such the approach suggested here is a random effect approach to account for the statistical heterogeneity and it will be explained in more detail in the next section.

According to one referee who reviewed this paper the above procedure cannot be applied for the following reasons: (a) The funnel plot is a mixture of studies some having probability of successes, *p*, statistically equal to 0.5 and others (approx. 30% of them) not equal to 0.5 and therefore as they do not have the same standard error they cannot be subjected to the same treatment. (b) The addition of data points in the void areas of a funnel plot, aiming to make it symmetric, will not increase the variance but decrease it. The reasons why I still believe the considerations presented here are applicable, against the opinion of the referee, will be discussed in the following.

Argument (a): I am not estimating standard errors for any of those studies. I simply place over the N=f(p) funnel plot the associated theoretical curves of a random process, for which p=q=0.5 for comparison purposes. This comparison of experimental data with theory is the standard approach in statistical decision theory, yet no statistical inferences are made here on the data. Finally, for $p \neq q \neq 0.5$, the standard error of proportions would indeed be different than in case of p=q=0.5, (see equation A-1 in appendix), but different in the sense of being now tighter, at a given study size, N, not broader as this funnel plot displays! In other words, all those other 30% of funnel plot data points for which p>0.5 or even p<0.5, should clustered tighter than chance expects about the funnel plot peak centered away from 0.5. This present funnel plot distribution of experimental points is exactly the opposite, and so much broader than would be for p=q=0.5! In the previous sections, furthermore, considerations regarding the origin of data as a mixture of different statistical distributions were not able to account for the present shape of the funnel plot.

(b) The unbiased funnel plot should be symmetric about its peak with no areas void of data points. The asymmetric funnel plot, due to those 'holes' of data in its distribution, has obviously a tighter scatter of data about their (shifted/biased) average than would have its hypothetical unbiased replica. The variance of the latter is thus larger than that of the former funnel plot as long as both contain datasets comparable in size. For as long as the funnel plot has a shape broader than one would expect due to chance alone, the lack of areas void of data in their scatter plot will enhance statistical heterogeneity, simply because now the difference of data from each other would be enhanced and that manifests statistical heterogeneity. One could make a mistake thinking that the statistical heterogeneity of a symmetric/unbiased funnel plot is lower than that of an identical one with 'holes' on the basis of the variance, simply because removing data to make 'holes' in the scatter plot will make the size, N, smaller and may increase the estimated variance which has N in the denominator if the sample size is very small. Similarly, if data points are added in a biased funnel plot to fill in the 'holes' and make it symmetric the scatter of data about their (new) average will increase and may either decrease or increase the variance depending on whether the proportion of added data is large compared to the size of database or not. The variance is not always decreasing with the addition of data points in void areas of the funnel plot.

HOW WELL DOES THE MICRO-PK GLUING IDEA ACCOUNT FOR THE FUNNEL PLOT EVIDENCE?

As explained above the funnel plot suggests that the effect of direct human intention on a true random process has broadened the scatter of proportion of successes by about five standard errors at each study size, yet leave the statistical average of proportions at the value expected by chance. It does not shift the chance average by a fixed number of standard errors. It does not add into the random process a fixed proportion of successes. It does not constitute a biasing mechanism, acting separately from the random process, either shifting the mean in the direction of intention, or even against it. Therefore, the initial hypothesis that micro-PK shifts the statistical average of a random process cannot hold up. There is an alternative hypothesis of a micro-PK mechanism, the statistical balancing and psi-induced persistence (gluing) (Pallikari, 2003) to try out. We shall show that the micro-PK gluing in combination with the statistical balancing approach describes a mechanism which accounts for the evidence provided by the funnel plot, its asymmetry and its convergence to chance expectation.

Success in a micro-PK test is the agreement between what is wished for, i.e. mentally perceived, and the actual outcome of a trial. Let us represent it by the digit 1. Failure, on the other hand, is the disagreement between the actual outcome of one trial and the outcome mentally perceived or wished for and can be represented by the digit 0. Just by chance sequences of the same bit can occur (runs) and their number, a_m, can be estimated in a sequence of N trials, where m is the length of the run. It is assumed that the bits are generated at the same probability p=q and that N>>m, (i.e. the number of trials is large enough). Then a_m and its variance s_m^2 are given by (Von Mises, 1964)

$$a_m = N\left(\frac{1}{2}\right)^{m+1} \qquad\qquad s_m^2 = a_m \tag{3}$$

The gluing in micro-PK suggests that there is a mechanism which acts on the generation of random events like 'psychic glue' to weakly sustain the length of any run with no preference on a specific type of bit (Pallikari, 2003). It adds momentum to the generation of a run, regardless of whether it is a run of successes or a run of failures and prolongs it by the same degree of persistence. This idea has arisen from a fractal analysis of micro-PK data (Pallikari, 1999). It was reported in that publication that even if there was no mean shift a low degree of persistence was present in runs of numbers either bigger, or smaller, than the theoretically expected average. Numbers above the average could be represented by 1s and those below the average by 0s and so the same nomenclature in terms of bit sequences can be obtained.

Since the micro-PK does not distinguish between successes or failures, zeros or ones, it will make the number of runs of either bit to increase equally. In the long run, that is at large enough study sizes consisting of many trials, the number of zeros and ones will balance out. As N becomes smaller the situation is quite different since the length of the run and the size of study N become more and more comparable. At large enough N a good balance between successes and failures will be obtained and pi will converge to 0.5, something that is harder to sustain at small N. To illustrate this, let us try to cover a narrow corridor with adjacent black and white tiles of length, m, and width equal to the width of the corridor, N. The longer the available length of each tile, m, compared to the length of corridor, N, the more distinct the disproportion in the number of black and white tiles used. Similarly, the shorter the study becomes in relation to the length of a run, the higher the imbalance between proportion of successes and failures and the stronger their scatter about 0.5. This is to be expected just by chance, but if the psi gluing increases the length of a run it also increases the scatter of pi (about 0.5), the associated funnel plot gets broader and the balance between the number of two types of tiles requires more trials, N, to manifest itself. The shorter their length the better the balance between the number of black and white tiles used. Similarly the larger the study compared to the length of run (that micro-PK can only moderately increase) the better a balance between successes and failures will be obtained and their proportion will eventually converge to 0.5. We have to assume that the length of black and white tiles increases by the same degree, that the persistence affects successes and failures equally, to justify the fact that in the long run balance is achieved and no mean shift is observed (Jahn, 2000). It has been estimated that a statistical balance between the two types of bits while micro-PK is underway will require about one million bits to be established (Pallikari, 2003) while in its absence about 10,000 should be enough. Since statistical balancing is obtained sooner or later one would expect large enough studies or comprehensive databases to display no mean shift. But that is clearly the case in early tests with card guessing (Rhine, 1969), in which the decline effect was first observed, or within the PEAR database (Jahn, 1987) in which the whole of experimental, control and calibration data exhibited no mean shift. In Rhine's card guessing tests the number of successes was balancing the number of misses and the overall outcome was displaying null mean shift. So, the gluing hypothesis is supported by evidence other than the micro-PK funnel plot.

One can try to describe this behaviour in a reverse manner. Since in the long run there will be no mean shift, should the data be pooled into two groups, (experimental and control) and the one set exhibits a significant statistical deviation from chance then the other will also exhibit a significant deviation from chance yet in the opposite direction. This observation has been termed 'statistical balancing' (Pallikari-Viras, 1997, Pallikari, 1998). Such statistical balancing is always underway in any random process but, if the psi gluing is under way, it is reached at a much slower rate with study size N. The micro-PK gluing will broaden the scatter of proportion of successes in micro-PK tests at all study sizes, which will always average to 0.5.

PREDICTING THE SCORE IN A MICRO-PK TEST

Let us expand the previous discussion a step further and investigate if the micro-PK can be predicted. If the micro-PK, or generally the mind-matter, is a mechanism introducing a state of entanglement between the mental and remote random physical process, then the person exhibiting psychic ability is skilled to manipulate it, engage to it or disengage from it, at will. In this entanglement the state of the remote system becomes the perceptive state of the psychic's brain. Since the micro-PK gluing applies indistinguishably to successes and failures, this 'manipulation' cannot unfortunately allow for the prediction of the score of a micro-PK test. We can neither predict the exact proportion of successes in a study, nor whether the score will be a success or failure.

Table 1: Approximate boundaries of the scatter of proportion of successes in a micro-PK test for study sizes up to 100,000.

Number of trials	Proportion of successes	
	Lower bound	Upper bound
100	0.23	0.77
1,000	0.41	0.59
10,000	0.47	0.53
100,000	0.49	0.51

Fortunately on the basis of statistical considerations, independent of shape of the present funnel plot, the range of pi values in the presence of micro-PK broadening mechanism can be predicted and the current micro-PK funnel plot appears to conform to it. In a study generating N trials in total the score pi will be bound within the limits set by equation 4, for N>30:

$$\boxed{pi = \begin{cases} 0.5 - \dfrac{2.7}{\sqrt{N}} \\[2ex] 0.5 + \dfrac{2.7}{\sqrt{N}} \end{cases}} \tag{4}$$

The prerequisite N>30 results from the need (a) the proportions of individual micro-PK tests to be normally distributed and (b) for equation 4 to vary between 0<pi<1 while z˜5 to envelope all data points (see appendix). Equation 4 suggests that the value of z in equation 1 be equal to 5.4. Thus an upper and lower bound for the effect size of micro-PK test can be estimated, some examples of which are given in table 1 for a range of study sizes. Exploiting equation 4 can obviously find tremendous technological applications.

Similar funnel plots can also treat the other types of psi phenomena, telepathy and remote viewing, provided careful and thorough meta-analyses are performed on their databases and predictions are made.

CAN PSI-GLUING BE ALSO UNDERWAY IN TELEPATHY AND REMOTE VIEWING PROCESSES?

Having demonstrated the importance of the gluing mechanism to explain the micro-PK process, it would make sense to assume that it relates to the fundamental nature of psi phenomena, such as telepathy and remote viewing. Let us now explore this possibility. While success in micro-PK means the outcome of a trial to agree with pre-stated intention, in telepathy (e.g. in a Ganzfeld test) it means the mental images between the two brains to convey the same meaning. In the case of remote viewing, success is equivalent to a match between the mental and the target images. Let us assume that the mental image of the recipient in a telepathy or remote viewing test consists of a very large number of pixels of information, same as in a hologram. Therefore, the telepathy test can be described as a sequence of agreements and disagreements (successes and failures) in information between the pixels of the mental images in the two brains. In remote viewing, the external to the brain image will represent an event or place.

We can now apply the psi-gluing concept into telepathy and remote viewing. On the assumption that the mind-matter entanglement sustains the run length of successes and failures, persistence of successes will provide larger chunks of meaningful, telepathically conveyed, information regarding the remote target and assist the viewer in its accurate identification. A persistence of disagreement between the pixels of two mental images, or the mental and the remote physical images, on the other hand, is equally likely, while it introduces noise with its irrelevant and nonsensical information. The process of logical inference of the brain can, in this case, boost only the persistence of agreement between pixels to complete a partially accurate image. Logical inference becomes inactive in building mental images in the case of persistence of nonsensical pixels of information, as it creates blurred, noisy mental pictures. The telepathically conveyed information to the brain, a mixture of larger chunks of meaningful information and smaller chunks of nonsensical noise, will thus create the basic psi-induced mental image. Since it contains a fair amount of noise, the telepathic image may occasionally be reminiscent, but not quite identical, to that of the remote target. For instance, the handle of an umbrella, as the target, may be perceived as the tail of a cat. The logical inference in the perceptive brain states of the remote viewer takes over and enhances the accuracy of the perceived image, acting only on the agreement (successes) between pixels. Without psi-gluing the meaningful pixels of the perceived images the telepathic process would be incomplete and inaccurate. Consequently, it is the combination of the psi gluing and the logical inference in the brain's perceptive states that characterize the telepathic or remote viewing skill.

CONCLUSIONS

It is shown that the recent meta-analysis of micro-PK data sheds light onto the nature of the micro-PK mechanism. This mechanism cannot constitute a real separate process different from the random process, adding either a fixed proportion of successes to it, or shifting the statistical average by a fixed number of standard errors. It is a mechanism that broadens the scatter of proportion of successes by a normal distribution of effect magnitudes. The hypothesis that micro-PK shifts the theoretically expected statistical average of outcomes generated by a random process must be abandoned.

A psi-gluing process that sustains the randomly generated runs of successes and failures can adequately account for the evidence in the shape of the micro-PK funnel plot, its convergence to 0.5 and its broadened width. Considering many alternatives the present asymmetry of the funnel plot is more likely to be due to unreported or missed studies. The micro-PK gluing and the statistical balancing processes explain why it is, unfortunately, not possible to predict whether the score of a micro-PK test will turn out to statistically be success, failure or simply chance. Fortunately, what can be predicted is the boundary of the broadened range of values that the proportion of successes can take in a micro-PK test as compared to those from a control test. Tremendous technological applications may result from this awareness.

It is therefore proposed that the definition of micro-PK will have to change assuming that it is the mental modulation of the statistics of a random process the cause of the funnel plot broadening. Micro-PK cannot manifest itself just by a shift of average in the direction of intention since it can also trigger it in the opposite direction especially at small studies and can even yield a null mean shift. In fact the null mean shift will be more and more likely as the number of trials in a study becomes larger and larger. More generally, micro-PK will be exhibited in an unbiased collection of micro-PK studies by a funnel plot which is symmetric centered about chance, 0.5, yet broadened in relation to a funnel plot of an identical collection of control scores. After about 1,000,000 of trials collected the experimental and control scores will not further differ statistically. In this sense, micro-PK will be expected to generate scores not only beyond chance especially at small N, within well-defined boundaries, but also within chance. To test the presence of micro-PK fast random number generators will have to generate datasets, e.g. of sizes from $N=50$ up to $N=10^9$, both at micro-PK and control conditions and the corresponding funnel plots be compared.

ACKNOWLEDGEMENT

I would like to warmly thank Dr Stefan Schmidt, the program chairperson of the 2004 PA convention, for his comments based on the reviews of his referees from which my manuscript has greatly benefited both in content and appearance.

REFERENCES

Bösch, H. (2002-3) Private communication.

Cooper, H. and. Hedges, LV eds. (1994). *The handbook of research synthesis*, Russell Sage Foundation, New York, pp. 302-305.

Deeks, J.J., Altman, DG., Bradburn, MJ. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger, Davey Smith, Altman (Eds.*) Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books

Deeks, J.J., Higgins, J. P. T., Altman, D. G., editors. Analysing and presenting results. In: Alderson, P., Green, S., Higgins, J., editors. *Cochrane Reviewers' Handbook 4.2.2* [updated December 2003]; Section 8. http://www.cochrane.org/resources/handbook/hbook.htm (accessed 31st January 2004).

Glossary (2001), *Journal of Parapsychology, 65* (4), 429-32.

Ibison, M., Jeffers, S. (1998). A double-slit diffraction experiment to investigate claims of consciousness related anomalies. *Journal of Scientific Exploration, 12*:4, 543-550.

Jahn, R. G., Dunne, B. J. (1987). *Margins of Reality: The Role of Consciousness in the Physical World* (Orlando: Hartcourt Brace Jovanovich).

Jahn, R., Mischo, J., Vaitl, D. et al (2000). Mind/Machine interaction consortium: PortREG replication experiments. *Journal of Scientific Exploration, 14* (4), 499-555.

Jeffers, S., Sloan, J. (1992). A low-light-level diffraction experiment for anomalies research. *Journal of Scientific Exploration, 6*(4), 333-352.

Jennions, M. D., Møller, A. P. Hunt, J. (2004). Meta-analysis can "fail": reply to Kotiaho and Tomkins. *OIKOS, 104*:1, 191-193

Pallikari, F. (1998). On the balancing effect hypothesis, in *Research in Parapsychology 1993*, ed. N. Zingrone (pp. 102-3). Metuchen: NJ: Scarecrow Press.

Pallikari, F. (2003). Must the 'Magic' of Psychokinesis Hinder Precise Scientific Measurement? *Journal of Consciousness Studies, 10*: (1-7), 199-219.

Pallikari, F. Boller, E. (1999). A rescaled range analysis of random events. *Journal of Scientific Exploration, 13*, 25-40.

Pallikari-Viras, F. (1997) Further evidence for a statistical balancing in probabilistic systems influenced by the anomalous effect of conscious intention, *Journal of the Society for Psychical Research, 62*: (849), 114-137

Palmer, A. R. (2002). Chimpanzee Right-Handedness Reconsidered: Evaluating the Evidence With Funnel Plots. *American Journal of Physical Anthropology, 118*: 191-199.

Radin, D., Nelson, R. (2002). Meta-analysis of mind-matter interaction experiments: 1959-2000, in *Spiritual Healing, Energy, Medicine and Intentionality: Research and Clinical Implications*, ed. W.B Jonas Edinburgh: Harcourt Health Sciences.

Scargle, J. D. (2000). Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration, 14* (1), 91-106.

Spiegel, R. M. (1961). *Theory and Problems of Statistics.* Schaum Publishing Co. New York

Steinkamp, F., Boller, E., Bösch, H. (2002). 'Experiments examining the possibility of human intention interacting with random number generators: A preliminary meta-analysis'. In C. Watt (ed.) *Proceedings of the Parapsychological Association, 45th annual convention*, pp. 256-272.

Von Mises, R. (1964). *Mathematical Theory of Probability and Statistics*, Ed: Hilda Geiringer, Academic Press: New York and London.

Address for correspondence: Fotini Pallikari,
Faculty of Physics, Solid State Physics Department,
University of Athens, Panepistimioupoli Zographou,
157 84 Athens, Greece.
E-mail: fpallik@phys.uoa.gr

APPENDIX

When testing the effect of a certain influence onto a system repeatedly we come up with a set of outcomes in the form of successes and failures to observe an influence. We then test the null hypothesis: that there is no effect, so the successes do not carry different weight than the failures, their respective probabilities are equal, p=q= ½, and the average proportion is equal to ½. If the actual average differs significantly from ½ this would imply that the probabilities, p & q are not equal, but one usually starts with the null hypothesis. Even if the Steinkamp et al meta-analysis contains some studies with p≠0.5, it has adapted the pi measure to be used in the funnel plot even for unequal probabilities.

The proportions pi estimated for each study of size N determine the sampling distribution of proportion of successes with mean μ_p and standard deviation σ_p which for an infinite population in the general case are (Spiegel, 1961, page142):

$$\mu_p = p, \quad \sigma_p = \sqrt{\frac{pq}{N}} \tag{A-1}$$

and for the case of equality between probabilities they are:

$$\mu_p = 0.5, \quad \sigma_p = \frac{0.5}{\sqrt{N}} \tag{A-2}$$

The z value is defined, for the case of equal probabilities, by:

$$z = \frac{pi - \mu_p}{\sigma_p} = \frac{pi - 0.5}{0.5/\sqrt{N}} \tag{A-3}$$

The estimated z value can be further used to support or reject the null hypothesis according to criteria of statistical significance, but in this work we are not going to apply such criteria. We note, however, that a value z=1.96 implies (since for N>30 the sampling distribution is very closely normally distributed, while the population is binomially distributed) that we expect the 95% of pi values to be found, transforming equation (A-3), in the region bounded by the curves obeying

$$N = 1.96^2 \frac{0.5^2}{(pi - 0.5)^2} = \frac{0.9604}{(pi - 0.5)^2} \tag{A-4}$$

The relationship $N = f(pi)$ from equations A-4 & 1 can be plotted on top of the scattered data of the funnel plot and it will be represented by two envelope curves, as in figure 1. Given that the sampling distribution of proportions, pi, is normally distributed, the 95% of micro-PK successes must lie between the two envelope curves (for z=1.96) and this clearly is not observed in the particular funnel plot of micro-PK experimental scores. Yet, this expectation regarding the data scatter is fulfilled in the case of the associated control studies (Bösch, 2002) as well as for the computer replication data obtained by the author.

Since the scatter of experimental data exceeds the width between the two (z=1.96) envelope curves per study size, one must obviously increase this width in order to envelope them all. By trial and error it is estimated that the appropriate envelope curves correspond to equation (A-5), which is equivalent to a z value close to 5 without intending to draw any statistical conclusions of this fact at this point. It is an empirical estimation as far as the funnel plot allows but not an arbitrary one. The equation (A-4) then becomes

$$N = \frac{6.25}{(pi - 0.5)^2} \tag{A-5}$$

Alternatively equation A-3 can be written as (see also equ. 1)

$$pi = 0.5 \pm z\frac{0.5}{\sqrt{N}} \tag{A-6}$$

Since the proportion of successes is limited in the region of values $0 \le pi \le 1$ and the sampling distribution is very closely normally distributed for $N \ge 30$ one can investigate the implications of these prerequisites. (a) For $pi \ge 0$ equation A-6 becomes: $0.5 \pm z\frac{0.5}{\sqrt{30}} \ge 0 \Rightarrow 0.5 \pm z\frac{0.5}{5.4} \ge 0$ which implies that $|z| \le 5.4$, z_{max}=5.4

(b) For $pi \le 1$ equation A-6 becomes: $0.5 \pm z\frac{0.5}{\sqrt{30}} \le 1 \Rightarrow 1 \pm \frac{z}{5.4} \le 2 \Rightarrow \pm\frac{z}{5.4} \le 1$ and 5.4 ≥ $|z|$ as found before.

This estimation is empirical, since it is not derived directly from a theory but from experimental data and it does not obligate any statistical inferences made on the meta-analysis result. It only indicates the extent of funnel plot broadening due to micro-PK and allows for predictions.

SYNCHRONICITY AND PSI: HOW ARE THEY RELATED?

John Palmer
University Hospital Zürich

ABSTRACT

Carl Jung's interest in parapsychology was kindled by personal psychic experiences, especially an anomalous banging sound emanating from a bookcase while he was in the company of Sigmund Freud. He also had an interest in Spiritualism and communicated extensively with J. B. Rhine about experimental parapsychology. He credits parapsychology with influencing the development of his theory of synchronicity. The purpose of this paper is to compare and contrast synchronicity experiences with psi experiences. Jung defined synchronicity as "the occurrence of a *meaningful* coincidence in *time*" (italics added). Re time, Jung generally viewed the elements of a synchronicity as simultaneous, but he did allow for a temporal relationship akin to precognition. With respect to meaning, Jung insisted that to be synchronicities, coincidences must reflect the activation of "archetypes", primal themes of good an evil inherited from our ancestors, and be of psychodynamic relevance to the experiencer. Psi experiences are not so restricted. Braude has criticized Jung for erroneously claiming that events can have inherent meaning. This problem can be overcome by saying that events differ in their capacity to evoke meaning in a person. Which events have this capacity must be defined by consensus among Jungian scholars. Jung postulated that one element of a synchronicity must be an experience and the other an objective event; this would seem to exclude pure telepathic experiences as examples of synchronicity. Synchronicity is unpredictable in the same sense that ESP card-guessing subjects cannot reliably predict whether any given response would be a hit. Jung performed an astrology experiment, finding that married couples shared astrological signs relevant to marriage more frequently than chance, but said that such statistical significance was not necessary; parapsychologists would not be so generous. Synchronicity cannot be the consequence of volition, which would seem to preclude most laboratory psi results; however, one can solve this problem by postulating that volition only *accompanies* psi in experiments. Mansfield, who sees no overlap between psi and synchronicity, subsumes psi under Jung's principal of "general acausal orderliness". At the level of physical theory, Jung subsumed synchronicity under quantum mechanics as interpreted by Pauli, and also saw correspondences to relativity theory. Compared to theories of psi, synchronicity theory comes closed to Stanford's conformance behavior model, but Stanford insists conformance is causal. The controversy over whether it makes sense to consider synchronicity acausal can be viewed as a semantic issue related to different definitions of causality being implicitly adopted by Jung and his critics. So long as synchronicity can be considered to some degree nomothetic, it can be tested empirically. At the physical level, testing can be considered isomorphic to tests of related psi theories such as the observational theories. Assuming a list of archetypes and evoking circumstances can be gleaned from Jung's writings, the theory can be tested psychologically by comparing test results under synchronicity conducive and non-conducive conditions. Jung suggested that evoking trance-like states and a sense of the miraculous can evoke archetypes. One might also explore manipulation of the archetypal content of ESP targets.

INTRODUCTION

Because parapsychologists populate a "borderland" area of science, we have always been grateful when eminent scientists and scholars from other fields reveal a positive interest in the subject, even more so if they make constructive intellectual contributions to it. One such eminent scholar active during the early to mid-twentieth century was the Swiss psychiatrist Carl Jung. As noted in an excellent review by Roderick Main (1997), Jung's interest in parapsychology was kindled by personal psychic experiences. Perhaps the most famous of these involved Jung's mentor, Sigmund Freud, who at the time was skeptical of psychic phenomena. During an argument about parapsychology, a loud sound, like an explosion, emanated from Freud's bookcase (Jung, 1963). When Freud denied that this anomalous event was paranormal, Jung predicted that it would immediately reoccur, and right on cue, it did. However, when the same thing happened to Freud again when Jung was not present, Freud discounted the original experience as being paranormal (McGuire, 1994). It was not until many years later that he expressed acceptance of psychic phenomena. His continuing difference with Jung on this matter has been cited as one of the reasons for their breakup (Main, 1997).

Jung also was very interested in Spiritualist mediumship and attended a number of seances, including one with the famous Rudi Schneider during which anomalous physical effects were observed (Charet, 1993). Although he found mediumship fascinating from a psychological point of view, by virtue of the dissociated states in which the mediums entered, he did not at the time consider the attendant phenomena to necessarily be paranormal (Jung, 1902/1969). Later, however, he experienced anomalous phenomena in a guest house in Britain where he had spent several nights alone and, as he subsequently learned, was reported to be haunted (Jung, 1950/1969). One class of phenomena, dripping water, he was unable to account for normally. Jung credits his experiences with psychic phenomena as helping him develop his psychological theory, especially the concept of synchronicity (Main, 1997).

Jung became acquainted with experimental parapsychology through the writings of J. B. Rhine, who was responsible for moving the field from its emphasis on spontaneous ("real-world") experiences, mediumship, and the question of post-mortem survival to controlled and statistically evaluated laboratory experiments with ordinary people, such as college students (Mauskopf & McVaugh, 1980). Over a period of about 25 years beginning in 1934, Jung and Rhine exchanged a series of letters that were uniformly friendly in tone and indicated mutual admiration. Jung accepted Rhine's evidence for the validity of psychic processes and also considered these data as support for his theory of synchronicity. Jung even proposed to Rhine an experiment on retrocognition (psychic knowledge of past events), which Rhine tactfully rejected, pointing out that retrocognition cannot be differentiated operationally from contemporary psychic awareness. Rhine nonetheless welcomed Jung's support and repeatedly urged him to publish his thoughts in English. On the other hand, the two men agreed to disagree on whether the mechanism for psi was causal or acausal, with Jung, of course, defending acausality. However, Rhine's defense of causality was far from dogmatic.

Synchronicity

Among Jung's many theoretical ideas, those of most direct relevance to parapsychology concern his concept of synchronicity. His most extensive presentation of these ideas was an essay entitled "Synchronicity: An acausal connecting principle" (Jung, 1952/1969). A shorter, but according to Main (1997), a more lucid presentation of the theory was published one year earlier (Jung, 1951/1969).

A famous example of a synchronistic event occurred during a therapy session with one of Jung's patients:

> My example concerns a young woman patient who, in spite of efforts made on both sides, proved to be psychologically inaccessible. The difficulty lay in the fact that she always knew better about everything....After several fruitless attempts to sweeten her rationalism with a somewhat more human understanding, I had to confine myself to the hope that something unexpected and irrational would turn up....I was sitting opposite her one day, with my back to the window, listening to her flow of rhetoric. She had had an impressive dream the night before, in which someone had given her a golden scarab — a costly piece of jewellery. While she was still telling me this dream, I heard something behind me gently tapping on the window. I turned round and saw that it was a fairly large flying insect that was knocking against the window-pane....I opened the window immediately and caught the insect in the air as it flew in. It was a scarabaeid beetle...whose gold-green colour most nearly resembles that of a golden scarab. I handed the beetle to my patient with the words, "Here is your scarab." This experience punctured the desired hole in her rationalism and broke the ice of her intellectual resistance. The treatment could now be continued with satisfactory results. (Jung, 1951/1969, pp. 525-526)

The best simple definition of synchronistic events I could find was appended to an English translation of Jung's 1952 essay: "the occurrence of *a meaningful coincidence in time*" (Jung, 1955, p. 144). There are two key words in this definition — *time* and *meaning* — each of which needs elaboration.

Time. With reference to time, Jung appears to insist that the corresponding events must be simultaneous. This creates some conceptual difficulties, because he clearly wants to include correspondences that parapsychologists would label as precognitive, that is, events in which the relevant external event followed the experience in time. His solution can be found in the addendum to his 1952 essay:

> [Synchronicity] can take three forms:
> a The coincidence of a certain psychic content with a corresponding objective process which is perceived to take place simultaneously.
> b The coincidence of a subjective psychic state with a phantasm (dream or vision) which later turns out to be a more or less faithful reflection of a "synchronistic", objective event that took place more or less simultaneously, but at a distance.
> c The same, except that the event perceived takes place in the future and is represented in the present only by a phantasm that corresponds to it.
> Whereas in the first case an objective event coincides with a subjective content, the synchronicity in the other two cases can only be verified subsequently, though the synchronistic event as such is formed by the coincidence of a neutral psychic state with a phantasm... (Jung, 1955, pp. 144-145)

Main (1997) points out a number of problems with this formulation, one of which is identifying what constitutes the "neutral objective state" in particular instances. I agree with Main's criticisms, but I will not discuss them further because they are not relevant to the main issue at hand. The point to note is that the correspondence between the subjective experience and the relevant objective event (whether or not the latter is, technically speaking, a component of the synchronicity) need *not* be simultaneous.

A second important point to be noted from Jung's description above is that in all three instances one of the two corresponding events is a subjective experience or process. In fact, if (b) and (c) are to be taken literally, in these cases one of the events must be a subjective *experience* exclusively, although I am not inclined to take him that literally here. Also, by "objective" I take him to mean external, which would not allow *both* components of the synchronicity to be subjective experiences or processes.

Meaning. Jung's definition of synchronicity is much more theory driven than parapsychologists' definition of psi, and this becomes most evident when we consider the second key term in Jung's definition of synchronicity, meaning. In Jung's theory, synchronistic correspondences are expressions of, and derive their meaning from, *archetypes*. The latter, which Jung equates to instincts (Jung, 1936/1969, p. 44), are primal themes, such as good and evil, that we inherited from our ancestors and are universally present in all human beings, where they reside in what Jung calls the *collective unconscious*. To count as synchronistic, a correspondence must be meaningful in Jung's sense of the term.

Stephen Braude (1979) has argued that it is nonsensical to suggest that either correspondences or the events that comprise them are inherently meaningful or non-meaningful. Meaning, Braude maintains, is imposed on events by the persons who observe them. I agree with Braude's point, but I believe its damage to Jung's theory can be mitigated through a subtle change in wording. Instead of saying that objective events are meaningful (or non-meaningful), we can say that they differ in their *capacity to evoke a sense of meaning or meaningfulness in the observer*. (This capacity in turn is governed by the archetypes, which are the ultimate source of meaning, as noted above.)

But how can we decide whether a correspondence has the capacity to evoke meaning? Isn't that a subjective judgment, and, if so, doesn't that belie the objectivity of the capacity? I think the answer is no, but to defend that answer I must appeal to the principle of *consensual validation*. By this I mean that the capacity (or lack of capacity) to evoke meaning is validated by a group of qualified evaluators, which in this case would consist of persons versed in Jungian thought. (I would want to add as a further condition that the evaluators agree on *what* the meaning is.)

There is nothing ad hoc about my appeal to consensual validation, because this is how we make decisions about objective reality all the time. I believe that the desk I am sitting at is real because I am confident that anyone else confronting the desk will have the same visual, tactile, and (if they tap on it) auditory experiences that I do. If others did not report these experiences, I would begin to question my judgment that the desk is really there. Such differences in sensory experience are precisely what often happens when someone claims to see a ghost. The resulting lack of full consensual validation is one of the main reasons apparitions are often interpreted by parapsychologists as hallucinatory.

Clearly, only a few among the almost infinite number of possible correspondences between psychic states and external events that one encounters in a lifetime would strike anyone as meaningful, even using the more mundane definition of meaningful. This subsample can then provide a meaningful (again in the mundane sense) set of correspondences at which to direct theoretical attention.

In fact, Braude's objection applies just as much to parapsychology as to synchronicity theory. Consider guessing through a deck of ESP cards. Out of all the possible patterns of matches between the guess and the target card, the only ones parapsychologists consider meaningful are those in which the matched symbols on a certain proportion of trials are identical, or bear some other pre-specified kind of "non-random" relationship to one another. This criterion of meaningfulness is imposed on the data by the parapsychologist, but in a way that is agreed upon by the qualified members of the community.

Psi and Synchronicity

During the remainder of this paper, I intend to address the relationship between synchronicity and psi by asking the following two questions:

(1) To what extent are the correspondences labeled by Jung as sychronistic the same as the psychic events studied by parapsychologists?

(2) If there is some overlap in the correspondences addressed by Jung and parapsychologists, how do the interpretations of these overlapping correspondences by the respective theorists compare?

THE CORRESPONDENCES

Subjectivity

Both Jung and the parapsychologists maintain that at least one element of the relevant correspondences in nature must be subjective. Although parapsychologists allow that both elements of the correspondence may be subjective, Jung says that one element of a synchronicity must be objective. Thus *telepathic* correspondences in parapsychology would *not* seem to qualify as synchronicities.

Time

Parapsychologists do not require that the corresponding events be simultaneous. Jung says the same about synchronicities; as discussed above, his decision to reformulate such correspondences as completely intra-psychic is a theoretical move not relevant to defining the correspondences at the purely descriptive level. This is also why I do not think this move allows us to consider telepathic exchanges as synchronistic.

Predictability and Statistical Significance

Physicist Victor Mansfield (1996) points out that synchronistic correspondences are "irregular", by which he apparently means that one cannot predict with certainty whether or not they will occur. This conclusion applies just as much to psi as to synchronicities. It is most obviously true for spontaneous cases of psi; no one can predict when such an experience might occur, including the person who has the experience. Although some repeatability has been demonstrated in laboratory psi experiments, it is statistical in nature: significant results appear more frequently than expected by chance, but not every time. Moreover, we are not very good at predicting when it will occur and when it will not. In other words, there is no "repeatability on demand" in parapsychology. It is not entirely clear whether Jung would accept statistical repeatability as a possible characteristic of synchronicities, but his discussions of Rhine's research lead me to believe that he would.

The one experiment on synchronicity that Jung reported himself deserves mention because it illustrates a curious conception he had about the role of statistics in demonstrating synchronicity. The experiment involved astrology (Jung, 1952/1969). Jung collected 483 pairs of horoscopes from married couples. He arbitrarily divided the horoscopes into three groups of unequal sizes. He noted retrospectively that the most frequently occurring conjunction of astrological signs for the paired horoscopes in each of the three samples, although different from one another, were ones predicted by astrologers to be characteristic of married couples. If we can assume that these are the *only* three of the 50 possible conjunctions that meet this criterion (this point is not clear from the report), then this outcome is highly significant statistically, even if one takes into account that the result was not hypothesized in advance. However, Jung is almost apologetic about this significance, and he clearly implies that the correspondence would have been meaningful, that is, a genuine synchronicity, even if the outcome had *not* been significant. (Note that it still might meet the standard of consensual validation discussed above.) This is a marked departure from the standards operative in parapsychology, which relies heavily on statistics to validate its evidence. Parapsychologists adopt this stance because they do not want to make the mistake of affirming correspondences that are not really there. They would never accept as real a correspondence that demonstrably falls within the bounds of chance (or the null hypothesis). Even using my more inclusive definition of psi, only correspondences that *might* be of non-chance character (such as good spontaneous cases) would fall in this category.

Volition

Mansfield (1996, p. 12) proposes that "any parapsychological effect that depends upon volition, upon the subject's intent, must differ from synchronicity". It is true that the examples Jung uses consistently conform to Mansfield's proposal. The one notable exception seems to be Jung's classification of Rhine's ESP results as synchronistic; Rhine's card-guessing subjects were consciously intending to demonstrate ESP during the test. Mansfield considers this endorsement to represent a contradiction in Jung's thinking. I disagree. Whereas Jung clearly states that synchronicities are not *caused* by a subject's conscious intent (Jung, 1960/1975, p. 541), this does not rule out the possibility that synchronicities might *accompany* conscious intent. This is what I think Jung would say happens in card guessing tests. Something about the situation in a card-guessing task can, under the right circumstances, activate an archetype and thus lead to a true synchronicity manifesting in the test itself. More on this in the next section.

Meaning

Jung maintains that synchronicities are meaningful in a very specific sense, namely that their meaning reflects the activation of an archetype. This kind of meaning literally becomes part of the definition of a synchronicity. As noted above, parapsychologists do not define psi in anything close to such a theoretical manner. Those who maintain that at least some psychic correspondences are paranormal rarely make any theoretical commitment beyond that.

To what extent might ordinary psi correspondences be susceptible to archetypal interpretation? Mansfield (1995, 1996) criticizes Jung for citing Rhine's card guessing results as examples of synchronicity because he (Mansfield) does not find an archetypal interpretation of them plausible. However, I suspect Jung would disagree. Specifically, Jung interprets the *decline effect* in Rhine's data — the tendency of scores to drop from highly positive to chance during a lengthy session — as evidence that the early success was accompanied by emotion or affectivity, which dissipated as the test proceeded. This emotion is seen as indirect evidence that an archetype had been activated (Jung, 1951/1969, 1960/1975). Moreover, although the standard ESP card symbols — star, circle, cross, square, wavy lines — were originally selected as being emotionally neutral, they would seem to have archetypal significance. For example, the cross has meaning in Christianity and some magical traditions, and the circle has significance in many Eastern traditions — the mandala, for example (Jung, 1969). Nonetheless, card guessing results are probably the hardest of all psychic correspondences to square with Jung's synchronicity theory. Spontaneous cases are invariably much richer and offer much more opportunity for psychodynamic interpretations. Some survey data suggest that a disproportionately high percentage of spontaneous psi experiences concern emotional themes (e.g., Schouten, 1979), although random survey data suggest otherwise (Palmer, 1979) and the preponderance of emotional cases in Schouten's surveys could be a reporting artifact.

Mansfield's (1996) positive proposal is that psi be considered a special case of what Jung called *general acausal orderedness* (GAO). Von Franz (1992, p. 267) defines GAO as a "regular omnipresent just-so-ness, such as for instance, the specific speed of light, the quantization of energy, the time-rate of radioactive decay, or any other constant in nature". This definition suggests to me that Jung intended GAO to refer to first-order physical laws of nature that cannot be reduced to some other law or cause. Jung (1952/1969) considered synchronicities to be a special case of GAO.

Mansfield's suggestion is congenial to parapsychologists, because we too consider psi to reflect the operation of an underlying natural law or laws. We resent it very much when psi is labeled as "supernatural" or "miracles". In fact, this lawfulness is inherent in the very definition of parapsychology. Note, however, that the classification of psi as GAO does not preclude it also having some overlap with synchronicity.

Conclusion

Based on the previous discussion, it can be concluded that there is indeed some overlap between synchronistic and psychic correspondences, but the overlap is not total. There are some synchronistic correspondences that are not psychic, and there are some psychic correspondences that are not synchronistic. The degree of overlap is difficult to estimate, and it is influenced by what assumptions are made about the two types of correspondences.

As noted previously, insofar as we are concerned about the *subject matter* of parapsychology and Jungian analysis, it is appropriate to include events that *might be* synchronistic or psychic (in the sense of paranormal) even though they are not demonstrably either one. This allows the inclusion of many spontaneous cases, thereby dramatically increasing the number of events of interest. It would also be expected to increase the degree of overlap between synchronistic and psychic correspondences as defined in this more objective manner.

Foremost among anomalous correspondences that are *psychic but not synchronistic* are those in which at least one element is psychological but for which an archetypal interpretation is sufficiently implausible to be ruled out a priori. The size of this population depends a great deal on how liberally the archetype hypothesis is applied. If archetypes are considered to be evoked by a wide range of circumstances, this population could become quite small. Second, cases of telepathy would seem to fall in this category. If we were to accept Mansfield's application of his volition restriction, most, but not all, laboratory evidence of psi would qualify. However, as I argued previously, I don't believe this application is justified so long as it is reasonable to suppose that accompanying psychological events might trigger a correspondence.

Foremost among anomalous correspondences that are *synchronistic but not psychic* are those involving at least one psychological element and for which an archetypal interpretation is plausible, but that demonstrably do *not* meet the criterion of statistical significance. An example here might be *some* so-called *Barnum statements* that psychics often give in their readings when nothing "good" seems to be coming through. These are statements that are so vague and general that they would apply to most people. "You have had pain in your life" is an example of a Barnum statement.

Finally, there exists in principle a class of anomalous correspondences that are neither synchronistic nor psychic: those that lack a mental component. I say "in principle" because I am hard pressed to think of a real example of this sort that has actually been put forth as paranormal (although I suspect readers better versed in these sorts of things than I am will not have this difficulty). A hypothetical example of an astrological nature would be the concordance of some pattern of stellar conditions and some purely physical event such as an earthquake.

THE INTERPRETATIONS

Now it is time to consider the interpretations offered by Jung and parapsychologists respectively to explain the overlapping correspondences that fall in the domain of both. Let me begin by reinforcing the point that although the distinction between events and their interpretation is useful, it is not absolute. This is because the definitions of both psychic and synchronistic correspondences are to some degree theory driven, although more so in the case of synchronicities.

Jung's Theory of Synchronicity

Jung's theory of synchronicity has both psychological and physical aspects. The psychological aspect is closely identified with the concept of the archetype. It is archetypes that "provide the shared meaning by virtue of which two events are considered to be in a relationship of synchronicity" (Main, 1997, p. 20). Archetypes thus provide the psychodynamic element to the theory. A more detailed analysis of their role is not necessary for the purposes of this paper.

The physical aspects of the theory draw heavily on modern (post-Newtonian) physics, especially quantum mechanics (QM). Jung's ideas about this were heavily influenced by his long association with the eminent physicist Wolfgang Pauli (Atmanspacher & Primas, 1997), who accepted Jung's synchronicity theory.

Jung's approach to the relationship between synchronicity and QM is to draw analogies between them, or, more precisely, between the processes of the psyche and the behavior of matter and energy at the subatomic level. "...we are concerned first and foremost", he writes, "to establish certain analogies, and no more than that" (Jung, 1947/1969, p. 234). Two such analogies are worthy of note:

1) Both radioactive decay and synchronicity are *probabilistic*, and hence (according to Jung) acausal (Jung, 1952, p. 512).

2) The *complementarity* of waves and particles in QM is analogous to the broader complementarity of causality and acausality in Jung`s theory (Jung, 1952, p. 514).

Jung also refers to the fundamental principle in QM commonly labeled as *collapse of the state vector*. He writes, "Physics has demonstrated...that in the realm of atomic magnitudes an *observer* [italics added] is postulated in objective reality....This means that a subjective element attaches to the physicist's world picture..." (Jung, 1947/1969, p. 230). In this passage, Jung appears to adopt what is in fact a minority viewpoint among physicists, namely that the consciousness of an observer is what collapses the state vector, thereby causing an undetermined state of a wave or particle to be determined. (The majority view is that the physical process of measurement causes the collapse.) Jung thus sees modern physics as supporting his view that there is a pervasive psychological component to objective reality.

Finally, Jung draws from Einstein's theory of relativity the notion that space and time are "relative" (Main, 1997), and he sees this relativity as particularly characteristic of the realm of the psyche. He also draws heavily on this principle to explain psychic phenomena such as clairvoyance.

Acausality. A fundamental principle of Jung's theory of synchronicity is that synchronistic correspondences are *acausal*, as opposed to causal. In fact, acausality goes well beyond synchronicity in Jung's thinking. Recall that general acausal orderedness (GAO) encompasses many of the fundamental laws of nature, and Jung explicitly characterizes QM as acausal.

Jung's proposal of acausality has been roundly criticized by parapsychologists (Beloff, 1977; Braude, 1979). These two critics both note that Jung uses the term causality in the narrow sense of Newtonian physics. This is the "billiard ball" kind of causality that requires the direct action of a "cause" upon an "effect". However, most philosophers (and I suspect most other people nowadays) use the term in the broader, Humean sense of a *contingency* between two events. As Beloff (1977, p. 577) puts it: "Let A and B be two different events, then if it *is* the case that B would not have occurred *but* for A, i.e., if A was necessary for the occurrence of B, we are justified in calling A the cause of B regardless of the temporal, spatial, material, or energetic relationships that may hold as between these two events". Thus, the controversy centers on how one prefers to use the term cause.

Jung seems to agree that synchronicities are causal in this broader, Humean sense when he writes, "But as soon as [one] perceives the archetypal background he is tempted to trace the mutual assimilation of independent psychic and physical processes back to a (causal) effect of the archetype, and thus to overlook the fact that they are merely *contingent* [italics added]" (Jung 1952/1969, p. 516). "Contingent" means causal in the Humean sense. In this sense, synchronicities are causal virtually by definition.

However, I find indications in Jung's writings that synchronicities are causal in a somewhat narrower sense of cause than the Humean. Consider, for example, when Jung (1952/1969, p. 481) writes, "The *effective…agents* [italics added] in the unconscious are the archetypes". It is true that the relationship between A and B, the two elements of a synchronicity, is merely contingent; there is no direct action of one upon the other. However, this does not rule out a third factor, C, which *does* have a direct action on A and B, in such a way that they come to coincide. C, of course, is the archetype. One need not assume there is any transmission or transfer of energy, so synchronicities may not be causal in the *strongest* Newtonian sense either.

I am aware that at other places Jung states quite forcefully that the role of the archetypes should not be seen as causal in the somewhat narrower sense; a good example is the quote from two paragraphs above. My point is that Jung at times writes about archetypes as if they are causal agents, even though when the matter is at the focus of his attention he denies it.

A possible explanation of Jung's apparently conflicted attitude is suggested by the following passage: "We are, however, driven to some such assumption if we are not to regress to a *magical causality* and ascribe to the psyche a power that far exceeds its empirical range of action" (Jung 1952/1969, p. 483).[1] Note the use of the word *regress*. Later on the same page, Jung writes in equally pejorative tones, "the *primitive* [italics added] mentality has always explained synchronicity as magical causality right down to our own day…" Could it be that Jung was afraid his ideas might become associated with ones that are considered highly objectionable by his broader culture, namely, magic? [2] (I am referring here not to stage magic but to what I will call natural magic).

Such a concern would be very understandable to many parapsychologists. A prominent critic of parapsychology not too long ago wrote a book entitled *Parapsychology: Science or Magic?* in which he attempted to denigrate parapsychology in part by linking it to popular notions of natural magic (Alcock, 1981). This rhetorical device can be effective insofar as magic evokes in the reader's mind associations such as witchcraft and Satanism. Although we should all strive to resist efforts to associate our theories to the excess baggage of popular occultism, the fact remains that the kinds of processes both parapsychologists and Jungians talk about, whether they be seen as causal or acausal, bear some resemblance to the underlying mechanisms of natural magic. It is true that Jung depersonalizes synchronicities by attributing them to the collective unconscious rather than the personal unconscious, and this is where the difference from magic is perhaps most evident. But putting the archetypes in the collective unconscious in no way precludes them from acting causally.

Theories of Psi

I divide theories of psi into two categories that I call *transmission models* and *correspondence models*. Transmission models assume that information in the form of a signal is carried from a source to a receiver over a channel. As a rule, these models propose the transfer of some sort of energy, either known or hypothetical, but that is not a necessary characteristic. Correspondence models are those that do not assume such transmission of information or energy, at least not in the classical, Newtonian sense.

Examples of a transmission model include the *ELF models*, which propose that psi information is carried from source to receiver by extra-low frequency electromagnetic waves (e.g., Persinger, 1979). Other theorists have proposed that the information is carried by subatomic particles with desirable properties, such as *neutrinos* and *tachyons* (e.g., Dobbs, 1967; Ruderfer, 1980). The hypothesized tachyons, for example, move

[1] Jung's apparent understanding of magical causality as being causal in a strong sense might be incorrect. In a dictionary of occult terms, Bletzer (1986, p. 366) lists as her primary definition of "magic causality": "the end result of two events happening together with a very significant meaning to the person and yet neither event is the cause of, or related to, the happening."

[2] Mansfield (1996, p. 17) also writes rather disparagingly of magic, referring to it as an "unwelcome guest". Ironically, in his book (Mansfield, 1995) he speculates that political considerations may have led Jung to equate psychic phenomena with synchronicities.

faster than the speed of light, which is helpful in explaining precognition. Other theorists have postulated even more novel forms of energy, such as "bioenergy" (Moss, 1974) and the Chinese Qi (Zha & McConnell, 1991), that are seen to be especially relevant to paranormal healing.

Examples of correspondence models include Stanford's (1978) *Conformance Behavior (CB) Model*, which interprets micro-PK, for example, as the "conformance" of a random event generator (REG) to the needs, wishes, or "dispositions" of the subject. In order to explain ESP, the brain is considered to be an REG, as the random firing of neurons in the brain bears a superficial resemblance to an REG's production of noise. Thus, in ESP, it is the brain that conforms to the subject's disposition.

Another class of correspondence models are those that propose a collective or group mind (e.g., James, 1909/1960; Price, 1940). The main idea of these models is that at an unconscious level we are all in effect one mind, so that interacting with the minds (and the knowledge) of others is really no different from interacting with our own minds. What prevents us from doing so in practice is that our brains filter out information from others, perhaps to prevent our own minds from being flooded with data and to maintain our personal identities.

The third class of theories that I place in the correspondence category are the *observational theories* (OTs), which are based on quantum mechanics QM (Millar, 1978). These theories take off from the notion that it is our consciousness, not the physical process of measurement, that collapses the state vector. It is proposed that consciousness can bias the collapse so that the selected state corresponds to someone's intention or desire. In parapsychology, this process occurs retroactively at the time someone sees feedback of the psi test result.

Other physics-based correspondence models propose that psi functions in higher dimensions of space/time than the four we are used to (e.g., Rauscher, 1983). As a useful metaphor, imagine a world in which the populace only experiences two spatial dimensions — length and width, but not height. If someone in tune with the third dimension were to drop an object from the sky onto the flat plane, it would appear to the observer that the object had suddenly come out of nowhere — what parapsychologists call an *apport*, a paranormal event.

Synchronicity and Psi

It should be obvious that the transmission models are very "non-Jungian", so if we are to look for good comparisons to synchronicity theory we must look to the correspondence models. At first glance, the collective or group mind theories might appear to be a good fit because of their apparent similarity to Jung's collective unconscious, which provides the milieu for his archetypes. However, Jung's collective unconscious is not a group mind in the sense used by parapsychologists. It is rather a group of individual minds that have the same contents derived from our common ancestry. Only if one interprets the collective unconscious in this way does Jung's equating it to instincts or instinctual patterns make any sense. On the other hand, the group mind of the parapsychologists is essentially *one* mind. Also, and perhaps more importantly, the parapsychologists' group mind contains all our mental impressions, including those that Jung would assign to his individual unconscious.

A better fit to synchronicity theory are those correspondence models based on modern physics. Both the OTs and synchronicity theory use state vector collapse by an observer as at least a good analogy to what happens in anomalous correspondences. Although Jung did not carry the analogy as far as did the authors of the OTs (for example, the assumptions underlying retroactive PK), the basic mechanisms of the two classes of theory appear to be identical as far as they go. Likewise, the multidimensional space/time theories of the parapsychologists seem to be making the same point as Jung when he talked out about the relativity of space and time, but once again the parapsychologists have developed the idea more fully than did Jung.

In my view, the closest fit to Jung's theory of synchronicity is Stanford's CB model. Without doing too much injustice to Jung's thought, I think one could describe Jungian synchronicities as the brain of a person coming into conformance with an external event in order to fulfill the person's needs and disposition. Of course, Jung's theory makes additional stipulations that go beyond the CB model or any of the other parapsychological theories. However, Stanford (1978) considers conformance to be a causal process.

TESTING SYNCHRONICITY

Testing Jung's synchronicity is not a simple undertaking. Mansfield (2002) states that "synchronicity's lack of an adequate theoretical structure; its uncontrollability; and its reliance on subjectivity, feelings, and scientifically suspect terms such as meaning make it exceedingly difficult, if not impossible, to test scientifically" (p. 178) Although these are indeed complications, they are no more severe than those routinely faced by experimental personality and clinical psychologists. Any general, nomothetic principles one can glean from Jung's writings can serve as hypotheses to be tested by conventional scientific methods.

Tests of the physical aspects of Jung's theory of synchronicity are isomorphic with tests of those parapsychological models such as the OTs that assume an acausal (in the Newtonian sense) correspondence between psychological and physical events. In a broader vein, Pallikari, Boller, and Bösch (2000) argue that cases in which data from micro-PK experiments exhibit effects that seem to reflect a rearrangement of targets within a sequence rather than a change in the relative frequency of the targets require for their explanation a concept akin to Jungian synchronicity. Pellikari et al. (2000) report data indicative of such patterning (excessive runs of the same target), and data conforming to decision augmentation theory (May, Spottiswoode, Utts, & James, 1995) would also seem to meet this condition. Whether such data can be accounted for by force-like causality models is a subject of debate (Dobyns & Nelson, 1998; May Utts, & Spottiswoode, 1995).

To my mind, the most promising and exciting possibilities for testing Jung's synchronicity theory come from its psychological aspect. It is clear from Jung's writings that archetypes can be activated by environmental stimuli (e.g., Jung, 1960/1975, p. 541), which means that in principle they can be activated in a controlled laboratory setting. In fact, social psychologists do this sort of thing all the time when they instill attitudes, beliefs, or emotions in their subjects in order to test their effects on behavior. (Of course, such manipulations sometimes raise ethical issues about the treatment of subjects, and these must always be weighed in the balance.) Jung also states that "synchronistic phenomena can be evoked by putting subjects in an unconscious (trance) state" (Jung, 1947/1969, p. 231). There is evidence that inducing a mild altered state of consciousness in subjects facilitates ESP (Honorton, 1977).

As a first step, it would be necessary to cull from Jung's writings, or those of his followers, a list of the archetypes. It also would be necessary to identify what specific kinds of environmental stimuli are capable of activating each archetype. The dependent variable could be any psi task that could also be synchronistic. The experimental design would specify two conditions or groups: one in which an archetype is activated right before the psi task, and one in which it is not (but is comparable in other respects). The hypothesis would be that the group which has been given the activation treatment would manifest significantly better psi scores than the other group. Of course, Jung would surely insist that the activation would not work every time, but it would not have to. Only a statistical effect is required.

There are other hints from Jung's writings that could be of value in helping parapsychologists improve the results they get in the laboratory. ESP targets might be selected or created that are relevant to particular archetypal themes, on the theory that subjects might be especially drawn to such targets, particularly if they relate to a particular subject's needs regarding individuation. Although the standard ESP cards are relevant to archetypes (as noted previously), more powerful examples could surely be devised. In discussing Rhine's card guessing experiments, Jung wrote that "the experimental set-up is influenced by the expectation of a *miracle* (Jung, 1960/1975, p. 537; italics in original). Although I am not sure that was quite the case, Rhine did place strong emphasis on motivating the subject to take the test (Rhine & Pratt, 1957). Be that as it may, I am not aware of any modern psi researchers who stress the "miraculous" nature of psi to their subjects. Jung's theory suggests that it might help our results to do so, by creating a sense of the "numinous" that would contribute to activating an archetype. It's certainly worth a try.

These examples of how Jung's ideas about synchronicity might be of concrete benefit to parapsychology is a good note on which to end on. Parapsychologists and Jungians have for years been addressing similar, even if not identical, phenomena, with little in the way of mutual interaction. Perhaps if our two groups were a little better "synchronized", both would benefit.

References

Atmanspacher, H., & Primas, H. (1997). The hidden side of Wolfgang Pauli: An eminent physicist's extraordinary encounter with depth psychology. *Journal of Scientific Exploration*, 11, 369-386.

Beloff, J. (1977). Psi phenomena: Causal versus acausal interpretation. *Journal of the Society for Psychical Research*, 49, 573-582.

Bletzer, J. G. (1986). *The Donning international encyclopedic psychic dictionary*. West Chester, PA: Whitford Press.

Braude, S. E. (1979). *ESP and psychokinesis*. Philadelphia: Temple University Press.

Charet, F. X. (1993). *Spiritualism and the foundations of C. G. Jung's psychology*. Albany: State University of New York Press.

Dobbs, A. (1967). The feasibility of a physical theory of ESP. In J. Smythies (Ed.), *Science and ESP* (pp. 225-254). New York: Humanities Press.

Dobyns, Y. H., & Nelson, R. D. (1998). Empirical evidence against decision augmentation theory. *Journal of Scientific Exploration*, 12, 231-257.

Eisenbud, J. (1970). *Psi and psychoanalysis*. New York: Grune and Stratton.

Foster, A. A. (1940). Is ESP diametric? *Journal of Parapsychology*, 4, 325-328.

Honorton, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435-472). New York: Van Nostrand Reinhold.

James, W. (1909/1960). The final impressions of a psychical researcher. In G. Murphy & R. O. Ballou (Eds.), *William James on psychical research* (pp. 309-325). New York: The Viking Press.

Jung, C. G. (1902/1969). On the psychology and pathology of so-called occult phenomena. In R. F. C. Hull (Trans.), *Collected works: Vol. 1. Psychiatric studies* (2nd ed., pp. 3-88). Princeton, NJ: Princeton University Press.[3]

Jung, C. G. (1936/1969). The concept of the collective unconscious. In R. F. C. Hull (Trans.), *Collected works: Vol. 9. The archetypes and the collective unconscious* (2nd ed., pp. 42-53). Princeton, NJ: Princeton University Press.

Jung, C. G. (1947/1969). On the nature of the psyche. In R. F. C. Hull (Trans.), *Collected works: Vol. 8. The structure and dynamics of the psyche* (2nd ed., pp. 159-234). Princeton, NJ: Princeton University Press.[3]

Jung, C. G. (1950/1969). Foreword to Moser: "Spuk: Irrglaube oder wahrglaube?". In R. F. C. Hull (Trans.), *Collected works: Vol. 18. The symbolic life* (2nd ed., pp. 317-326). Princeton, NJ: Princeton University Press.[3]

Jung, C. G. (1951/1969). On synchronicity. In R. F. C. Hull (Trans.), *Collected works: Vol. 8. The structure and dynamics of the psyche* (2nd ed., pp. 520-531). Princeton, NJ: Princeton University Press. [3]

Jung, C. G. (1952/1969). Synchronicity: An acausal connecting principle. In R. F. C. Hull (Trans.), *Collected works: Vol. 8. The structure and dynamics of the psyche* (2nd ed., pp. 417-519). Princeton, NJ: Princeton University Press.

Jung, C. G. (1955). *Synchronicity*. London: Ark Paperbacks.

Jung, C. G. (1960/1975). Letter to A. D. Cornell. *Letters: Vol. 2* (pp. 537-543). Princeton, NJ: Princeton University Press. [3]

Jung, C. G. (1963). *Memories, dreams, reflections*. New York: Pantheon Books.[3]

Jung, C. G. (1969). *Collected works: Vol. 9. Psychology and religion* (2nd ed.; R. F. C. Hull, Trans.). Princeton, NJ: Princeton University Press.

[3] Relevant text is reproduced in Main (1997).

Main, R. (1997). *Jung on synchronicity and the paranormal*. London: Routledge.

Mansfield, V. (1995). *Synchronicity, science, and soul-making*. Chicago: Open Court.

Mansfield, V. (1996). Distinguishing synchronicity from the paranormal. Unpublished manuscript.[4]

Mansfield, V. (2002). *Head and heart*. Wheaton, IL: Quest Books.

Mauskopf, S. H., & McVaugh, M. R. (1980). *The elusive science*. Baltimore: The Johns Hopkins University Press.

May, E. C., Spottiswoode, S. J. P., Utts, J. M., & James, C. L. (1995). Applications of decision augmentation theory. *Journal of Parapsychology*, 59, 221-250.

May, E. C., Utts, J. M., & Spottiswoode, S. J. P. (1995). Decision augmentation theory: Applications to the random number generator database. *Journal of Scientific Exploration*, 9, 453-488.

McGuire, W. (Ed.) (1994). *The Freud/Jung letters*. (R. Manheim, Trans.). Princeton, NJ: Princeton University Press.

Millar, B. (1978). The Observational Theories: A primer. *European Journal of Parapsychology*, 2, 304-332.

Moss, T. (1974). *The probability of the impossible*. New York: Tarcher.

Pallikari, F., Boller, E., & Bösch, H. (2000). Jungian synchronicity sheds light on the micro-PK mechanism? *Proceedings of Presented Papers: The Parapsychological Association 43rd Annual Convention*, 210-222.

Palmer, J. (1979). A community mail survey of psychic experiences. *Journal of the American Society for Psychical Research*, 73, 221-251.

Palmer, J. (in press). *Dulling Occam's razor*. San Juan, PR: Puente Publications.

Persinger, M. A. (1979). ELF field mediation in spontaneous psi events: Direct information transfer or conditioned elicitation? In C. T. Tart, H. E. Puthoff, & R. Targ (Eds.), *Mind at large* (pp. 189-204). New York: Praeger.

Price, H. H. (1940). Some philosophical questions about telepathy and clairvoyance. *Philosophy*, 15, 363-385.

Rauscher, E. (1983). The physics of psi phenomena in space and time. *Psi Research*, 2, 93-120.

Rhine, J. B., & Pratt, J. G. (1957). *Parapsychology: Frontier science of the mind*. Springfield, IL: Charles Thomas.

Ruderfer, M. (1980). Neutrino theory of psi phenomena. In B. Shapin & L. Coly (Eds.), *Communication and parapsychology* (pp. 121-149). New York: Parapsychology Foundation.

Schouten, S. A. (1979). Analysis of spontaneous cases as reported in "Phantasms of the living". *European Journal of Parapsychology*, 2, 408-455.

Stanford, R. G. (1978). Toward reinterpreting psi events. *Journal of the American Society for Psychical Research*, 72, 197-214.

Von Franz, M. (1992). *Psyche and matter*. Boston: Shambhala.

Zha, L., & McConnell, T. (1991). Parapsychology in the People's Republic of China: 1979-1989. *Journal of the American Society for Psychical Research*, 85, 119-143.

Address for correspondence: John Palmer,
University Hospital Zürich, CH-8091 Zürich.
E-mail: john.palmer@usz.ch

[4] This paper is available at Dr. Mansfield's web site: http://lightlink.com/vic/mlv_paper.html

EXPERIMENTER EFFECTS IN LABORATORY TESTS OF ESP AND PK USING A COMMON PROTOCOL

Chris A. Roe[1], Russell Davey[1] & Paul Stevens[2]
[1]*Division of Psychology, University College Northampton*
[2]*Koestler Parapsychology Unit, University of Edinburgh*

ABSTRACT

This paper describes a fourth study designed to explore the relationship between ESP and PK performance by testing for both using a common protocol so as to control for expectancy effects and experimental artifacts. Following earlier work (Roe, Davey & Stevens, 2003, in press, unpub.) we were particularly concerned to look for possible experimenter effects upon performance. Forty participants completed a computer-based greyhound racing game. Races occurred in two blocks of 12. One block was presented as an ESP task and required participants to nominate which of the six greyhounds had won a race that the computer had already run silently. The program then replayed the race as feedback. The other block was presented as a PK task and required participants to 'will' a greyhound that was selected for them to run faster than its competitors. The greyhound's movements were determined in real time by an RNG for PK trials and using pseudorandom data for ESP trials. Within each block half the races were in fact ESP trials and half PK trials, presented in random order. For half the trials the second author served as experimenter as in previous studies, but for the other trials the first author fulfilled this role. After chatting with the participant, the experimenter completed an interaction questionnaire to evaluate the warmth, spontaneity and positivity of the briefing. Performance was non-significantly better than chance overall, and significantly better for disguised ESP trials. There were no significant relationships between performances in the four conditions, although the effect sizes were of a similar magnitude and direction to those found previously. Participants who had been briefed by the first author performed better than those briefed by the second author, in line with prediction overall; suggestively so overall ($p = .06$) and significantly so for Disguised ESP and PK ($p < .001$ and $p < .05$ respectively). A number of the interaction measures gave promising correlations with task performance, particularly the experimenter's confidence that the session would be successful ($rs = -.431$, $p = .007$). Further research into interaction effects is recommended.

INTRODUCTION (HEADING)

In recent work (Roe, Davey & Stevens, 2003, in press, unpub) we have been concerned to address the question of whether ESP and PK functioning are sufficiently distinct to merit separate terms. Earlier work that had considered this issue is difficult to interpret because the method of testing for ESP is typically quite different from that for PK so that any apparent differences in the preferred conditions of the phenomena may be artifactual (Schmeidler, 1988). We developed a new protocol using a computer game interface that allowed both phenomena to be tested for within a standardised context. In the game, RNG and pseudorandom data are sampled to determine the movements of six greyhounds from the left to the right of the screen, simulating a race. The program monitors progress and registers the order in which the dogs cross the finishing line. In the ESP condition a race had been run 'silently' so that the outcome was 'known' to the computer. Participants were informed that their task was simply to select one dog from among the six that they felt had performed best on that trial. They then watched a replay of the race and the result was confirmed. In the PK condition the race would be run in real time with the movements of their pre-selected greyhound determined by a random number generator (RNG). Participants were informed that their task was to attempt to influence the RNG and thus enable their greyhound to succeed. The program consisted of a block of 12 races that ostensibly were all testing for ESP and a further block of 12 testing for PK. However, half of the trials that appeared to be tests of ESP in fact were of PK and vice versa in order to differentiate between characteristics of the phenomenon and participants' expectancies concerning that phenomenon.

The results from our first three studies have been somewhat disappointing, with overall performance at chance levels for both ESP and PK trials, and for true and disguised trials. One potential contributory factor to poor performance that has not previously been considered by us is that variables associated with the experimenter may have had an inhibiting effect. White (1977, p. 273), for example, has noted that "the experimenter has been a neglected variable in parapsychological research ... [yet] ... there could hardly be a more significant area of investigation than the role of the experimenter". Rhine and Pratt (1957) have characterised the experimenter as having to be able to provide "the psychological conditions under which psi can operate" (p. 131, cited in White, 1977, p. 274), and Gardner Murphy (1949) suggested that there is no such thing as a gifted participant as such, but rather how well a participant scores on a psi task depends on the person who does the testing and the nature of the experimental conditions. These experienced researchers seem to share the view that the experimenter plays a crucial role in encouraging or inhibiting psi in the laboratory (see Smith, 2003, for an updated consideration of this issue).

Within parapsychology there seems to be a common belief that some experimenters are psi conducive, whereas others appear to be psi inhibitory (cf. Irwin, 1999; Smith, 2003). Indeed there is quite clear evidence of an experimenter effect leading to differential performance on psi tasks in circumstances where all other conditions (and even in some instances the participants themselves) are common to both. For example, Van Busschbach (1956) describes a study in which Rhea White and Margaret Anderson tested schoolchildren in different classes but of the same age and in the same schools. The latter oversaw above-chance scoring (p = .002) whereas the former's participants performed at chance, with the difference between experimenters significant (p = .02). Similarly, Nicol and Humphrey (1953) found striking differences in the results of the same participants when each of the authors served as experimenter using the same test under similar conditions, and Bednarz and Verrier (1969) reported a similar finding, albeit with overall negative scoring. Pratt and Price (1938) have also noted that they showed differential performance with the same participants.

One possible explanation of the effect is that the experimenter's personality, behaviour and enthusiasm may indirectly influence the results of a psi study by motivating participants or providing them with clues that provide further information about the nature of the study and about the experimenter's hopes or expectations. These demand characteristics may affect the subsequent behaviour of the participants and thus the results of the study itself (Harris & Rosenthal, 1985; Rosenthal 1966; see also White, 1977). For example, in terms of their communication skills, there is some evidence that, at least in a non-experimental context, psi-conducive and psi-inhibitory experimenters differ in how their body language is perceived by observers; psi-conducive experimenters might be considered to be more flexible, enthusiastic, friendly, likeable and warm (Schmeidler & Maher, 1981). In a more direct assessment, Honorton, Ramsay and Cabibbo (1975) had two experimenters who interacted either in a positive manner (friendly, casual, supportive) in which time was taken to establish rapport with the participant, or in a negative manner (abrupt, formal and unfriendly) in which they went quickly into the task. The positive treatment gave significantly higher scores than the negative. Woodruff and Dale (1950) were the first to include direct measures of the attitudes of experimenter and participant towards one another. Surprisingly, higher scores were associated with participants who gave lower ratings for experimenters. It should be noted however, that 'low' ratings were actually neutral rather than very negative, and may indicate that disinterest or antipathy is desirable. For Dale, there was a difference in participant scores such that those given higher liking ratings by her subsequently performed significantly better at the psi task. Other researchers to look at the degree of liking between experimenter and participant include Nash (1968) and Waldron (1959) though the results are less clear in these cases.

There is also evidence to suggest that the beliefs of the experimenter might influence the outcome of the study. Sharp and Clark (1937), using several experimenters who differed in their attitude towards psi, obtained results that seemed to relate experimenters' attitudes to participants' subsequent hit rates; Sharp (mean hit rate = 5.36, where MCE = 5.00) and Davidson (5.88) were positive towards the existence of psi, whereas Berger (4.86) was uncommitted, and Myers (4.30) was sceptical. However, typically for a quasi-experiment, it is difficult to know whether these experimenters may have differed from one another in other ways that might have influenced participant scoring, for example the more sceptical experimenters may have been less friendly, older, etc. Parker (1975) more directly manipulated expectancy among six 'experimenters' (student data collectors) so that they presented as strong believers or strong disbelievers. Although overall scoring was null, there was a significant difference in performance between the experimenter groups, even though they in fact were testing the same sender-receiver pairs.

Alternatively it may be that different outcomes reported by different experimenters are explicable in terms of some form of parapsychological experimenter effect — where the results are partially dependent on the experimenter's own psi ability (see Kennedy & Taddonio, 1976; Palmer, 1997; Schmeidler, 1997; White, 1976). However, on grounds of parsimony this study will primarily focus on psychosocial explanations.

Reflecting on the previous studies in this series, we should note that the researcher responsible for all interactions with participants was RD. Although RD has a Bachelor's degree in Psychology and has previously conducted a parapsychological study for his dissertation, he would nevertheless be considered a novice experimenter, whereas CR has been involved in a number of previous studies that have reported significant effects (e.g., Roe, 1996; Roe, Holt & Simmonds, 2003) and may simply be more practiced at engendering a psi conducive atmosphere. It could be that if a more experienced researcher had interacted with participants in these earlier studies then a different outcome might have occurred. Secondly, although involved in the later stages of design of these studies, RD was not involved at the project's inception and may not feel the same degree of 'ownership' of the project that CR would feel through having been responsible for the seed idea, conducting background literature research, writing funding proposals, and so on. On these grounds we suspected that participants tested by CR may perform significantly better than those tested by RD. We also planned to have the experimenter complete an assessment of key aspects of the interaction in the hope that we may be able to identify predictors of success. Finally, we intended to continue to take attitude and personality measures from participants in order to determine whether the preferred conditions for ESP and PK task performance were similar or distinct (see Roe et al., 2003, for a more detailed rationale for the inclusion of particular measures).

METHOD

Design

This study incorporated a 2x2 mixed design looking at the effects of task type (ESP versus PK) and experimenter (CR or RD). The former involved within-Ss comparisons, while the latter involved between-Ss comparisons. The primary outcome measure was pre-specified to be the weighted sum of ranks of finishing positions. We also intended to conduct exploratory correlational analyses to determine whether task performance in the four conditions covaried systematically with personality and attitude variables. All analyses were planned to be non-parametric.

Materials and apparatus

A participant information form (PIF) was constructed which asked about basic biographical and contact details. Of particular interest here, the PIF incorporated a version of Thalbourne and Delin's (1993) Australian Sheep Goat Scale (ASGS, adapted after Roe, 1998); the Kiersey Temperament Sorter (Keirsey & Bates, 1978) — a variant of the Myers Briggs Type Indicator; and both forms of Spielberger's (1983) State-Trait anxiety inventory. The PIF is a generic form that also includes various other questions (e.g., about hypnagogic/hypnopompic experiences) that were not planned to be a focus of this study. Copies of the PIF are available on request from the first author. A seven-item interaction questionnaire was devised specially for this study (a copy is included as an appendix). Items were included to gauge the experimenter's mood at the time of their interaction with the participant, as well as their rating of the warmth, spontaneity and positivity of the interaction (after Schmeidler & Maher, 1981).

A computer program was developed by PS that makes use of real-time true random versus pseudo-random data to move six greyhounds from the left to the right of the screen, simulating a race. The number of moves is determined by the output so that over successive iterations some greyhounds move closer to the finish than others. The program monitors progress and notes the order in which the dogs cross the finishing line. The program continues until all six dogs have completed the course. The participant's task is simply (in the ESP condition) to select a dog that they would like to own and that they thought had done well in the race, or (in the PK condition) to have their dog identified for them by the computer and for them to 'will' it to succeed. In either case, the participant 'wins' any prize-money awarded based on the dog's finishing position. Prize money is used as a simple weighted score based on finishing position (100 virtual pounds for first, £50 for second, £25 for third, no prize money for the other placings). After a series of races the participant amasses an amount of overall prize-money. The program consists of 24 races, taking approximately 12 minutes to complete. Races are run in two blocks of 12 races that ostensibly are either tests of ESP or PK. In fact within each block half the trials are of ESP and half of PK, presented in random order.

Participants

Forty people participated in this study, of whom 21 were males and 19 females, with a mean age of 31.4 years (*SD* = 13.6, *Median* = 24). Participants were drawn from an opportunity sample from the Northampton area. Relatively few were undergraduate students at University College Northampton.

Procedure

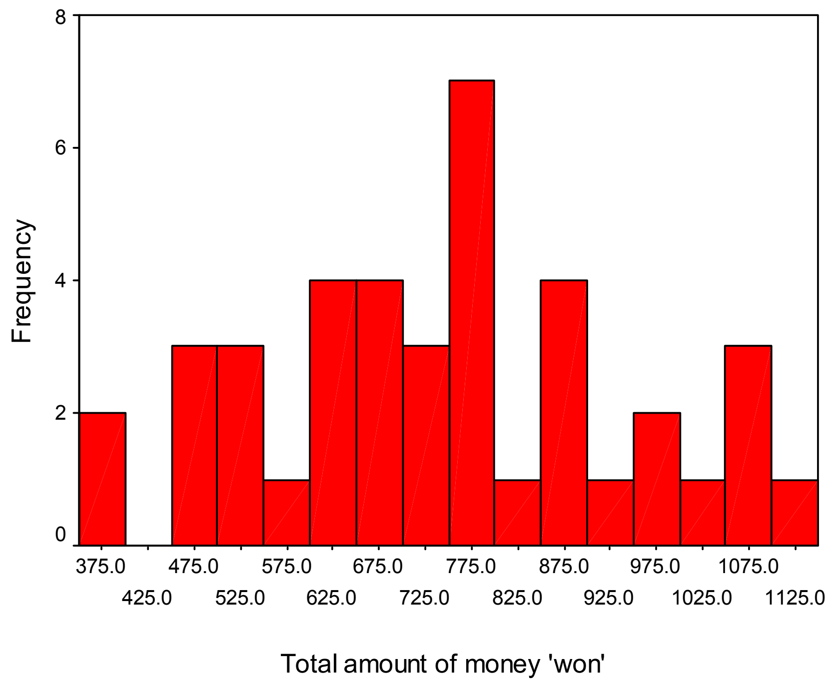
Prior to the session participants were given the PIF to take away and complete. They were greeted by the first or second author (CR or RD) who acted as experimenter. In some cases, participants had not completed the measure (e.g., if they had questions about certain items) in which case they were given time prior to their trial to complete the form. Participants next completed the state form of Spielberger's (1983) State-Trait anxiety inventory. The experimenter sought to make the participant feel relaxed and comfortable, engaging in casual conversation as well as explaining the nature of the task and answering any questions they might have.

Participants were then escorted by CR or RD into a research cubicle containing a PC with the program ready to begin. The program autoran and presented participants with a series of 24 races in two blocks of 12. One block was labelled as 'gambler' races and were ostensibly ESP trials. Here participants saw the onscreen briefing: "For the next 12 trials we'd like you to play the role of a gambler who has a free hand to choose which dog to select. In this session the races will already have been run by the computer but not yet have been played out. Your task is to use ESP to identify which of the 6 dogs won the race. Once you've made your choice you'll see a replay of the race on screen". Prior to each gambler race, participants were prompted to enter a number from 1 to 6 corresponding to their choice of dog for the forthcoming 'replay'. A second block was labelled as 'owner' races and consisted of ostensible PK trials. Here the onscreen briefing was: "For the next trials you will play the role of an owner whose greyhounds are entered in a series of races. Your dog will be pointed out at the beginning of each race, and its speed will be determined by a random number generator in the computer. Your task is to try to use PK to influence the RNG so that your preselected dog wins the race. You'll see the race in real time so you get feedback on how well you're doing". Prior to each owner race, participants were asked to press the space bar to start the race. All participants completed both blocks with the order of completion counterbalanced across participants. Within each block, half the trials were as given in the briefing (e.g., tested for ESP in the gambler block), but half were not (e.g., tested for PK in the gambler block) to gauge the effect of expectation on performance. The experimenter remained outside the research cubicle during trials but was available should assistance be required. During this time (i.e. before the outcome was known) the experimenter completed the interaction questionnaire. Once the participant had completed the first block of 12 races, the experimenter rejoined them to discuss how they were getting on and to engage in casual conversation so as to help reduce any fatigue and / or boredom and to reinforce the experimenter-participant relationship. They again waited outside while the participant completed the second block of trials. After the program had finished the experimenter debriefed participants, describing the nature of the four conditions within the task and explaining the need to disguise certain aspects of it. Given the mild deception involved, great pains were taken to ensure that participants were satisfied of the need for the study to be designed as it was and to be sure that they were happy for their data to be included in analysis. No participants asked to withdraw.

RESULTS

The planned outcome measure here is the finishing position of participants' greyhounds in computer races, but to get a sense of whether overall performance was above MCE we shall firstly consider the overall amount won by each participant. The greater the success at the task the greater the amount of prize money that will have been won. If chance alone is operating then a participant will typically have won prize money of £700. We can see from Figure 1 that the distribution of scores is relatively normal and peaks slightly to the right of the theoretical midpoint. In fact in this study the average prize money is nonsignificantly above this (M = £730.0, *SD* = £198.8; Wilcoxon Z = -.838, *p* = .402, 2-tailed).

Figure 1: Frequency histogram of prize money 'won' by participants



As previously, it was planned in advance to use sum of ranks for final finishing positions as the principal outcome measure. The distribution of ranks for each of the four conditions is given in Table 1. We can see that in terms of overall scoring, results in this study are somewhat better than previously. The overall sum of ranks for target dogs is below the MCE of 840 in two of the four conditions, significantly so in the case of disguised ESP. Performance in the two PK conditions closely approximates MCE, and for true ESP is non-significantly worse than chance expectation. Perhaps surprisingly, performance in the disguised ESP condition was significantly better than for the true ESP condition (reminiscent of Study 2), although there is no overall difference in performance across the conditions (Friedman's χ^2 = 5.52, *p* = .20). Unlike in previous studies there is no tendency for frequencies to increase as we move from first place through to sixth; correlating frequency against finishing position gives a nonsignificant negative Spearman's correlation (*rho* = -.314, *p* = .544).

Table 1: Sum of ranks for greyhound finishing position

Condition	*Finishing position*						*SOR*	*Z score*	*Effect size (r)*
	1	*2*	*3*	*4*	*5*	*6*			
MCE	40	40	40	40	40	40	840		
True PK	43	46	31	31	41	48	845	.170	.011
Disguised PK	44	34	43	37	44	38	837	-.094	-.006
True ESP	38	30	48	40	34	50	872	1.191	.077
Disguised ESP	46	50	42	34	37	31	779	-2.287	-.148
Total	171	160	164	142	156	167	3333	-.501	-.016

To consider whether similar patterns of performance across individuals are evident for ESP and PK conditions (either informed or disguised), we considered covariation of performance across the four conditions. Correlations of individual sum of ranks scores are given in Table 2.

Table 2: Spearman rho correlation coefficients (with p values in parentheses) for comparisons of individual performances in the four conditions

	true ESP trials	*Disguised ESP trials*	*true PK trials*
disguised PK trials	-.055	-.020	.095
	(.734)	(.902)	(.558)
true ESP trials		.043	.255
		(.793)	(.112)
disguised ESP trials			.114
			(.485)

We can see from this that none of the correlations comes close to statistical significance, indicating that performance in one condition cannot be predicted on the basis of performance in any of the other conditions. Given the relatively low power of this study we may want to consider the effect sizes themselves, and note points of similarity with previous studies in this series. As in two of the previous three studies, the largest effect size is the positive correlation between true ESP and true PK. None of the others has replicated consistently.

We were interested in this study to see whether there might be evidence of an experimenter effect. RD is a relative novice researcher and it was hypothesised that CR might be more successful, especially given that the study was originally designed by him and that he may feel more 'ownership' over it. We can see in Table 3 that the overall performance of CR's participants was suggestively superior to that for RD, and this difference was significant in the case of disguised ESP. Interestingly, CR's participants also showed slightly better performance in the other condition that included an element of deception, namely the disguised PK condition, whereas the reverse is true for true PK.

Table 3: Mean sums of ranks (and standard deviations) for informed and uninformed participants for the four conditions

	True ESP	*Disguised ESP*	*True PK*	*Disguised PK*	*Overall*
RD trials	21.80	21.35	20.75	21.95	85.85
	(4.76)	(3.66)	(4.01)	(3.86)	(8.98)
CR trials	21.80	17.60	21.50	19.90	80.80
	(4.29)	(3.33)	(3.91)	(4.53)	(9.07)
Mann Whitney Z	-.082	3.063	-.597	1.413	1.544
p (1-tail)	.234	.001	.137	.039	.062

To further understand the nature of any possible experimenter effect, the relationships between experimenter ratings on the interaction questionnaire and task performance are given in Table 4. We can see that there is a suggestive tendency for task performance to improve as the experimenter rates themselves as more relaxed. This seems particularly so for conditions that entailed deception. A similar pattern is evident with ratings of the orientation of the interaction, with greater positivity being associated with better performance, although these effects are smaller and clearly non-significant. It will be interesting to see if these patterns can be replicated. Other notable effects here include a significant relationship between current mood and performance, with more positive mood predicting better performance, significantly so in the case of disguised ESP. Perhaps most striking is the strong correlation between experimenter confidence of success and actual success, which suggests that on the basis of the experimenter's assessment of the interaction (or of qualities within the individual – though of course at this stage their individual differences measures had not been scored) they were able to predict how participants would fare in the psi task.

Table 4: Spearman correlations between task performance and experimenter's interaction ratings (probabilities in parentheses are two-tailed)

	True ESP	*Disguised ESP*	*True PK*	*Disguised PK*	*Overall*
How would you rate your current mood (1= negative, 7 = positive)	.089 (.596)	-.334 (.040)	-.182 (.274)	-.084 (.617)	-.187 (.262)
How relaxed are you (1=relaxed, 7 = tense)	.000 (.998)	.215 (.196)	.123 (.464)	.274 (.096)	.286 (.082)
How would you rate the warmth of the interaction? (1=very cold, 7=very warm)	.032 (.850)	.101 (.546)	-.049 (.771)	-.155 (.352)	-.065 (.698)
How would you rate the spontaneity of the interaction? (1=rehearsed, 7=spontaneous)	-.130 (.437)	.062 (.713)	-.185 (.266)	.318 (.052)	.081 (.630)
How would you rate the orientation of the interaction? (1=very negative, 7=very positive)	-.139 (.406)	-.116 (.486)	-.124 (.459)	-.154 (.355)	-.233 (.159)
How would you describe the quality of rapport with the P (1=extremely poor, 7=extremely good)	.116 (.486)	.065 (.697)	.014 (.935)	-.177 (.288)	.008 (.963)
How confident are you that the trial will be successful? (1=not at all confident, 7=extremely confident)	-.181 (.278)	-.253 (.125)	-.308 (.060)	-.206 (.216)	-.431 (.007)

Relationships between participants' personality and attitude scores and their performance at the tasks are given in Table 5. The patterns here are generally disappointing and none of the correlations comes close to significance.

Table 5: Spearman correlations between task performance and belief and personality variables (probabilities in parentheses are two-tailed)

	True ESP	*Disguised ESP*	*True PK*	*Disguised PK*
PK 'criterion 1'	-.202	.007	.033	.109
	(.237)	(.966)	(.847)	(.526)
ESP 'criterion 1'	-.107	.059	.060	.064
	(.542)	(.736)	(.734)	(.716)
Overall ASGS score	.015	.231	.008	.019
	(.933)	(.175)	(.963)	(.913)
ESP factor	.029	.127	.003	.031
	(.865)	(.459)	(.986)	(.859)
PK factor	-.088	.288	.021	-.029
	(.610)	(.088)	(.903)	(.866)
Survival factor	-.046	.345	-.190	-.127
	(.789)	(.039)	(.267)	(.461)
Prior experience	-.087	.150	-.122	-.236
	(.606)	(.370)	(.467)	(.154)
State anxiety on STAI	.095	.095	.023	-.060
	(.580)	(.583)	(.893)	(.728)
Trait anxiety on STAI	-.122	.203	-.007	-.110
	(.461)	(.216)	(.966)	(.504)
3-hour K index value	.097	-.055	.178	-.200
	(.550)	(.737)	(.271)	(.217)

Finally we attempted to replicate the claimed tendency for those who present as Feeling / Perceiving on MBTI measures to outperform those who present as Thinking or Judging types on GESP tasks (e.g., Honorton et al., 1990). The mean sums of ranks for Feeling-Perceiving and non-FP types are given in Table 6. Again, note that higher sums of ranks indicate worse performance at the task. We can see that these data fail to confirm previous suggestions of superior performance for FP types in ESP tasks (e.g., Honorton et al., 1990), with performance for the disguised ESP being in the predicted direction but not significantly so and performance in the true ESP condition being slightly better for nonFPs. For the PK tasks there was a slight superiority for FPs at the true PK task, but this falls well within chance expectation. However, there was a significant difference between FPs and nonFPs at the disguised PK task, with the latter faring better, contrary to expectation.

Table 6: Mean sums of ranks (and standard deviations) for FP and non-FP types for the four conditions

	True ESP	*Disguised ESP*	*True PK*	*Disguised PK*	*Overall*
Feeling-Perceiving	22.13	18.31	20.50	23.38	84.31
	(4.67)	(2.87)	(3.35)	(3.74)	(7.76)
Other	21.47	20.53	21.21	19.21	82.42
	(4.22)	(4.35)	(5.24)	(3.63)	(10.37)
Mann Whitney Z	-.233	1.368	.233	-3.028	-.415
p (2-tail)	.816	.171	.816	.002	.678

GENERAL DISCUSSION AND CONCLUSION

Although the overall performance of participants in this study is still not significantly better than chance, there is some room for optimism when we consider all the studies in this experimental series. The performance of participants has improved steadily across all four studies to date, with average prize money being respectively £648.1, £660.6, £715.0 and £730.0 (where MCE is £700), suggesting that the revisions made to the method at each stage might have had some effect. In this latest study the further improvement witnessed may be explicable in terms of an experimenter effect, since as hypothesised, Roe's participants fared better than Davey's, suggestively so overall and significantly so for disguised ESP trials. It would be very interesting to see how other researchers, particularly those who have been identified as being psi conducive (Smith, 2003) might fare using this method.

Another ground for optimism comes from the finding that a number of the experimenter-participant interaction items gave suggestive correlations with task performance, notably the experimenter's mood and level of relaxation as well as positivity of the interaction. Particularly interesting here was the quite strong correlation between experimenter confidence that a participant would do well and their subsequent performance, which could constitute the most important finding in this suite of studies. However, there was no encouragement to be derived from the correlations of task performance against spontaneity, warmth and degree of rapport, *contra* earlier suggestions (e.g., Honorton et al., 1975; Schmeidler & Maher, 1981).

Participant personality and attitudinal patterns again showed little evidence of association with ESP and PK task performance, which begs the question, if these factors are not predictive of success, then on what basis were the experimenters' confidence judgements being made? One speculative interpretation, given the tendency for experimenter-centred assessments to represent stronger predictors than participant-centred ones, would suggest that the experimenter must be more seriously considered to be the source of any psi. in considering the experimenter's contribution in future work, it would clearly be useful to have participants also complete the interaction questionnaire (we were fearful of overburdening them here after the other questionnaires that they had to complete) so that we might be better able to differentiate between situational factors and intra-experimenter factors.

ACKNOWLEDGEMENT

We would like to gratefully acknowledge the financial support of the Fundacao Bial (grant no. 58/00), which has enabled us to conduct this study.

REFERENCES

Bednarz, K., & Verrier, M. (1969). Role of experimenter in GESP tests. *Journal of Parapsychology*, 33, 159.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.

Honorton, C., Berger, R., Varvoglis, M., Quant, M., Derr, P., Schechter, E., & Ferrari, D. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology*, 54, 99-139.

Honorton, C., Ramsey, M., & Cabibbo, C. (1975). Experimenter effects in extrasensory perception. *Journal of the American Society for Psychical Research*, 69, 135-149.

Irwin, H.J. (1999). *An introduction to parapsychology (3rd ed.)*. Jefferson, NC: McFarland.

Keirsey, D., & Bates, M. (1978). *Please understand me: Character and temperament types*. Del Mar, CA: Prometheus Nemesis Books.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, 40, 1-33.

Murphy, G. I. (1949). Psychical research and human personality. *Proceedings of the Society for Psychical Research*, 48, 1-15.

Nash, C. B. (1968). Comparison of ESP run-score averages of groups liked and disliked by the experimenter. *Journal of the American Society for Psychical Research*, 62, 411-414.

Nicol, J. F., & Humphrey, B. M. (1953). The exploration of ESP and human personality. *Journal of the American Society for Psychical Research*, 47, 133-178.

Palmer, J. (1997). The challenge of experimenter psi. *European Journal of Parapsychology*, 13, 110-125.

Parker, A. (1975). A pilot study of the influence of experimenter expectancy on ESP scores. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds), *Research in parapsychology 1974*, (pp 42-44). Metuchen, NJ: Scarecrow Press.

Pratt, J. G., & Price, M. M. (1938). The experimenter-subject relationship in tests for ESP. *Journal of Parapsychology*, 2, 84-94.

Rhine, J. B., & Pratt, J. G. (1957). *Parapsychology: Frontier science of the mind*. Springfield, Ill.: Charles C. Thomas.

Roe, C. A. (1996). Clients influence in the selection of elements of a psychic reading. *Journal of Parapsychology*, 60, 43-70.

Roe C. A. (1998). Belief in the paranormal and attendance at psychic readings. *Journal of the American Society for Psychical Research*, 92, 25-51.

Roe, C. A., Davey, R., & Stevens, P. (2003). Are ESP and PK aspects of a unitary phenomenon? A preliminary test of the relationship between ESP and PK. *Journal of Parapsychology*.

Roe, C. A., Davey, R., & Stevens, P. (in press). Are ESP and PK aspects of a unitary phenomenon? A further test of the relationship between ESP and PK. *Journal of the Society for Psychical Research*.

Roe, C. A., Davey, R., & Stevens, P. (unpub.). Arousal and performance at ESP and PK tasks using a common protocol. Submitted to the *European Journal of Parapsychology*.

Roe, C. A., Holt, N., & Simmonds, C. A. (2003). Considering the sender as a PK agent in ganzfeld ESP studies. *Journal of Parapsychology*, 67, 129-145.

Rosenthal, R. (1966). *Experimenter effects in behavioural research*. New York: Appleton-Century-Crofts.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and mismatches*. Jefferson, NC: McFarland

Schmeidler, G. R. (1997). Psi-conducive experimenters and psi-permissive ones. *European Journal of Parapsychology*, 13, 83-94.

Schmeidler, G. R., & Maher, M. (1981). Judges responses to the nonverbal behaviour of psi-conducive and psi-inhibitory experimenters. *Journal of the American Society for Psychical Research*, 75, 241-257.

Sharp, V. & Clark, C. C. (1937). Group tests for extrasensory perception. *Journal of Parapsychology*, 1, 123-142.

Smith, M. D. (2003). The role of the experimenter in parapsychological research. *Journal of Consciousness Studies*, 10 (6-7), 69-84.

Spielberger, C. D. (1983). *State-trait anxiety inventory for adults: Form Y*. Redwood City, CA: Mind Garden, Inc.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research*, 59, 172-186.

Van Busschbach, J. G. (1956). An investigation of ESP between teacher and pupils in American schools. *Journal of Parapsychology*, 20, 71-80.

Waldron, S. (1959). Clairvoyance scores of sheep versus goats when subject's attitude toward the experimenter and the purpose of the experiments are manipulated. Abstract. *Journal of Parapsychology*, 23, 289.

White, R. A. (1976). The limits of experimenter influence on psi test results: Can any be set? *Journal of the American Society for Psychical Research*, 70, 333-369.

White, R. A. (1977). The influence of the experimenter motivation, attitudes and methods of handling subjects in psi test results (pp. 273-304). In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold.

Woodruff, J. L., & Dale, L. A. (1950). Subject and experimenter attitudes in relation to ESP scoring. *Journal of the American Society for Psychical Research*, 44, 87-112.

Address for correspondence: Chris A. Roe,
Division of Psychology, University College
Northampton, Park Campus, Boughton Green Road,
Northampton NN2 7AL, UK.
E-mail: chris.roe @northampton.ac.uk

Interaction questionnaire

This questionnaire asks about your thoughts and feelings after you have been briefed about the ESP experiment, and will hopefully give us some clues as to what factors are important for success. Your answers are completely confidential and will not at any time be seen by your experimenter. When you have completed the questionnaire please place it in the stamped addressed envelope so that it can be sent to an independent scientist for analysis.

Session number:	
Date of session:	
Relationship with participant	

1. How would you rate your current mood?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Negative … Positive

2. How do you feel at this moment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Relaxed … Tense

3. How would you rate the experimenter / participant interaction in terms of *warmth*, *spontaneity* and *positiveness*?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very Cold … Very Warm

☐ ☐ ☐ ☐ ☐ ☐ ☐

'Rehearsed' … Spontaneous

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very negative … Very Positive

4. How would you describe the quality of the rapport that you have with the participant?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Extremely Poor … Extremely Good

5. How confident are you that today's experiment will be a success?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Confident … Extremely Confident

A FURTHER CONSIDERATION OF THE SENDER AS A PK AGENT IN GANZFELD ESP STUDIES

Chris A. Roe & Nicola Holt

Centre for the Study of Anomalous Psychological Processes

University College Northampton

ABSTRACT

Recent work has been concerned to evaluate whether the sender plays any active role in successful ganzfeld GESP experiments (e.g., Roe, Holt & Simmonds, 2003). Earlier ganzfeld research that has considered the role of the sender has suggested a modest advantage for sender over no sender conditions despite weaknesses in general design that make any sender effect difficult to detect. The present study was a replication of Roe, Holt & Simmond's (2003) novel method for overcoming some of these criticisms by utilising a random event generator (REG) as a proxy receiver. During the sending period descriptive statements were 'selected' from among a pool of 768 to give a 20-item 'REG mentation' that may represent a more direct measure of any sender influence. Forty ganzfeld trials were conducted with novice sender-receiver pairs using a standard protocol apart from the introduction of the REG. Immediately prior to the sending period, the computer determined whether or not a sender would be shown the target or would engage in an alternative task. By this method 23 trials involved senders and 17 did not. Receivers registered a 25% overall hit rate, equal to mean chance expectation, with slightly better performance in sender trials than no sender trials (26.1% compared with 23.5%). The REG mentation was used by two independent judges to rank order the clips in each target set, and gave rise to overall hit rates of 30.6% and 16.7%. Considering sender and no sender trials separately, both independent judges gave higher hit rates for the sender trials than for the no sender trials (42.1% vesus 17.6% and 26.3% versus 5.9%), though these differences were not significant. There was no significant relationship between the ranks awarded by the receiver and those by the independent judge using REG mentations (▯ = -.174 and -.186), or indeed between the independent raters (▯ = .202). None of a number of individual differences measures significantly predicted receiver performance, and indeed none were suggestive ($p < .1$). There was a significant relationship between one judge's target rating and sender sending strategy, and a suggestive correlation with sender focusing method.

INTRODUCTION

Despite the relative success of clairvoyance designs in eliciting evidence for ESP when used with other protocols (see, e.g., reviews by Rhine, Pratt, Stuart, Smith & Greenwood, 1966, and Utts, 1996), it has been commonly assumed that a sender can make some positive contribution to the outcome of Ganzfeld studies. Relatively few Ganzfeld experiments have adopted a clairvoyance design except where the objective of the study was to compare sender and no sender conditions. Honorton (1995) found that of 73 Ganzfeld studies only 12 did not employ senders. His meta-analysis comparing sender and no sender experiments showed that those including senders generated better performance than those that did not, although the effect seemed to be confined to those experimenters who had used both conditions at some time. If it could be shown that the sender were unnecessary this would have practical advantages in that sessions would be easier to co-ordinate for only one participant at a time, and security would be more straightforward, since no person need know the identity of the target until after the participant's judgements had been recorded.

Eight previous Ganzfeld studies have directly compared sender and no sender conditions within the same study (see Roe, Sherwood & Holt, 2003, for a more detailed review), which together give a reasonably consistent picture. Two studies have reported scoring when a sender is aware of the target to be significantly better than when a sender is absent (Raburn & Manning, 1977; Sargent, Milton, Payne & Bennet, unpub, cited in Milton, 1988-9), and four describe a non-significant advantage for telepathy conditions compared with clairvoyance conditions (Dunne, Warnock & Bisaha, 1977; Kanthamani & Khilji, nd, described in Kanthamani & Palmer, 1993; Milton, 1988-9; and Williams, Roe, Upchurch, & Lawrence, 1994). Although

Morris, Dalton, Delanoy and Watt (1995) concluded that there were no significant differences overall, the reported target ranks give rise to *z* scores of 1.028 and .237 for their informed sender and uninformed no sender conditions respectively. Roe, Sherwood and Holt (2003) reported a slightly improved performance for sender trials compared with no sender trials in terms of direct hits but the reverse pattern when considering sums of ranks and *z*-scores of similarity ratings. Taken together, these findings may offer some encouragement for the suggestion that the sender serves some active role in a typical Ganzfeld ESP session.

These findings may be particularly encouraging given that the experimental manipulations of the IV here are rather gross, for example in not systematically taking into account the possible moderating effect of variables such as the sender-receiver relationship (see Honorton, 1985, but also Bem & Honorton, 1994). The designs also tend to assume that any sender effect will be readily apparent in the receiver's overall performance, despite this relationship being dependent upon the receiver not only being able to detect any sender mediated impressions, but also to accurately interpret them and to be able to discriminate them from internally generated 'noise' during judging. Recently we reported on an alternative method for gauging any potential sender effect that promised to circumvent such complications by replacing the receiver with an REG 'virtual receiver' that would generate a virtual mentation by randomly selecting statements from among an array of descriptors. Although a somewhat unusual procedure, a variant of this approach had previously met with some success (Roe, 1996).

Roe, Holt and Simmonds (2003) conducted an initial test within a ganzfeld setting. Here, 40 pairs of participants experienced a 'standard' autoganzfeld, but during the sending period a 'virtual' receiver in the form of an REG located in the receiver's room operated to generate random numbers corresponding to members of a pool of 768 statements. These had been coined to describe all of the clips in the whole pool (i.e. 8 descriptors for each of the 96 video clips). A virtual mentation was generated consisting of the 20 statements that were most frequently selected during the sending period. An independent judge rank ordered the four clips in each target set in the same manner as the actual receivers had done, but using correspondences derived from the virtual mentation as a basis for selection. Overall, the 'live' receiver selected the correct clip as the target on 14 occasions (35% hit rate where MCE is 25%), and by pre-planned sum-of-ranks analysis performed significantly better than chance expectation ($Z = 1.77$, $p = .038$). More interestingly in the present discussion, the ratings based on the REG-generated virtual mentations gave an encouraging 13 hits (32.5%), and a suggestive sum-of-ranks outcome ($Z = 1.48$, $p = .069$). While not wishing to over-interpret a non-significant outcome, it may be interesting to note that the associated effect size, r, of .234 is of a comparable magnitude to that reported in previous research that utilised an REG as a proxy psychic reader, and from which this study was adapted (Roe, 1996, $r = .257$). It is also orders of magnitude larger than typical effect sizes for other REG-based studies, which Radin and Nelson (1989) estimated at 3×10^{-4}, and may be more in keeping with those associated with the DMILS protocol, estimated at 0.33 (Braud & Schlitz, 1991)[1]. The outcome was regarded as sufficiently promising to warrant further investigation. In the present study we planned to compare performance of the REG on sender trials with performance on trials when there was no sender (or at least where a nominal sender was unaware of the target). In this case the mentation would presumably consist of random noise and might provide a more suitable control against which to evaluate performance in the experimental condition.

It was also found previously that whatever degree of success could be attributed to the REG condition, this was not a result of the REG actually selecting statements that were originally coined for the target clip more often than would be expected by chance. Rather, since the statements were purposely phrased so as to be somewhat vague or open to interpretation or which reflected themes that might be represented in a number of clips (e.g., "feeling really cold, freezing", "Floating gently down as if with a parachute", "Sounds of screeching or squealing") it seemed that the independent judge had been able to use statements originally

[1] However, more recently this has been calculated as 6×10^{-5} (Steinkamp, Boller & Bösch, 2002), derived from the weighted mean effect size, pi, of .50003. We are mindful, however, of the extreme heterogeneity of the results included in that analysis and the reported correlation with study size that may reflect a publication bias, and which leads us to be cautious in our interpretation (E. Boller, personal communication, February 2003).

intended for other clips to enable them to identify the target clip. Thus we could not discount the possibility that any above chance scoring was a function of the judge's ability, with the mentation simply providing a Rorschach-like medium for her own psi in what might amount to little more than an elaborate forced choice ESP test. It was not practicable to assess this possibility in Study 1, for example by involving a second independent judge, and so in this study we planned to recruit a second independent judge to rate the transcripts for comparison.

We have previously reported on a comparison of sender and no sender conditions in the ganzfeld study incorporating sender-receiver pairs described here. In this paper we report on a second aspect of that study in which an REG served as a virtual receiver during the normal session to generate an alternative mentation to be rated by independent judges.

Study aims

Two predictions were hypothesised using sum of target ranks as the primary outcome measure. Predictions given below are directional but alpha levels for other, exploratory analyses were conservatively kept as two-tailed:

1. REG mentations will allow an independent judge to identify the target clip to a greater degree than expected by chance
2. The independent judge's sum of ranks for sender trials will be lower than for no-sender trials

We also planned to conduct exploratory analyses considering covariation of REG and receiver performance with receiver and sender personality and attitude measures. These correlations were pre-specified as two-tailed.

METHOD

Design

The present study is the second of a series that is intended to systematically explore the utility of using an REG as a proxy receiver within a ganzfeld GESP protocol, and consists of a comparison of sender and no sender conditions. The dependent variable for planned analyses is the sum of target ranks awarded by the independent judge; the DV for exploratory analyses is the *z*-score of target ratings.

Participants

The sample used in this study is the same as that reported in Roe, Sherwood & Holt (2003). An opportunity sampling method was used to draw 40 pairs of participants (mean age of senders = 29.7 [range = 18 – 60], 14 males and 26 females; mean age of receivers = 28.0 [range = 18 – 60], 15 males and 25 females). These mainly consisted of friends and acquaintances of the experimenters, and staff and students at UCN, although attempts were made to recruit participants from the wider community using posters and media appeals. Each participant provided his or her own sender. Lab personnel did not serve as participants. The mean ASGS belief score for receivers in this sample was 44.5 (std dev = 10.4) and for senders was 46.6 (11.27). These figures are somewhat below the theoretical mean for the scale (mean = 54), suggesting that the sample is moderately sceptical. Three experimenters conducted trials in this study; Nicola Holt conducting 17 trials, Chris Roe 13 trials and Simon Sherwood[2] 10 trials.

[2] We are grateful to Dr Sherwood for his assistance with this project.

Apparatus and Materials

Details of the experimental suite, including distances between rooms have been described previously (Roe, Sherwood & Holt, 2003).

This study used an automated ganzfeld computer system developed by Dr Paul Stevens and written in Microsoft Visual Basic v5 that presented video material via the API for Media Player v7. Video clips are stored digitally as MPEG files, labelled 1a, 1b, 1c etc. Three separate monitors for the experimenter, sender and receiver are controlled by the experimenter PC via separate video cards, which prevents video leakage. Security measures within the program lock the experimenter out of system completely during a session, so that it is not possible to switch to another application or access the computer except by aborting the session. Audio signals are split into left and right channels for sender and experimenter/receiver so that it should not be possible for audio leakage to occur.

A revised target set was used for this study, consisting of 116 minute-long digital video clips arranged in 29 sets of 4. Copies of the target pool are available on CD or DVD from the first author on request. Ganzfeld randomisation is achieved using the Visual Basic pseudo-random algorithm (rnd), seeded using the timer at the start of the program (RANDOMIZE TIMER). Once the 'Start' button has been pressed, the computer first selects a target set, then selects one of the four clips within that set. The order of presentation of the four clips at judging is similarly randomised.

The descriptor pool from which the REG draws was the same as that for the original study (Roe, Holt & Simmonds, 2003), and consists of eight statements for each of the 96 clips to give a total pool of 768. These statements were coined by the authors to describe the original target set, but were intended to be accurate but not overly-specific (e.g., "a sense of flying or floating" rather than "a flock of birds flying overhead") so that they were more characteristic of the kinds of descriptions given during ganzfeld stimulation, and also so that they could in principle help identify targets from other sets. This had been the case in the previous study. Statements were selected during the sending period by taking one sample a second from an Orion RNG v1.1 attached to a Pericom 386 PC running under DOS. At the end of the sending period, the 20 statements that had been selected most often formed the mentation. In the event of any ties at the cut-off point of 20 then the first selected was automatically taken.

Materials

The Participant Information Form (PIF) is a 55-item measure that was constructed for general use with parapsychological research at University College Northampton and includes questions concerning biographical and contact details (11-items); religious and parapsychological background (5 items); computer experience (2 items); practice of mental/physical disciplines (2 items); belief in luck (2 items); clumsiness and punctuality (2 items); competitiveness (1 item); absorption (2 items); sleep and dreams (4 items); imagination and fantasy-proneness (3 items); creativity (2 items); and physical and mental health (1 item). The remaining items relate specifically to knowledge, belief and experience of anomalous phenomena including telepathy, clairvoyance, precognition, psychokinesis, 'communication with the dead' and out of body experiences (18 items); and hypnagogic/hypnopompic experience in a range of modalities (10 items). The form concludes with an open question inviting descriptions of personal anomalous sleep-related experiences. Copies of all in-house measures are available from the first author on request.

Participants also completed the short extraversion and neuroticism subscales of the EPQ-R (Eysenck, Eysenck, & Barrett, 1985). Each subscale has 24 items with a dichotomous yes/no response format. The 18-item Australian Sheep-Goat Scale (ASGS, Thalbourne & Delin, 1993), with a 5-point Likert scale ranging from strongly agree to strongly disagree, was also completed.

Post-ganzfeld measures included a Sender Strategy Questionnaire that asked about the type of sending strategies used, whether this was active or passive, holistic or atomistic, focused on target clip or on the receiver, realistic or associative, and continuous or episodic.

Procedure

Potential participants were sent an information sheet illustrated with photographs that described most aspects of the study. This provided a rationale for the ganzfeld paradigm, outlined the stages of the experimental procedure, focusing on the roles of the experimenter, sender and receiver. Participants were made fully aware that the sender might not be required for the Ganzfeld trial, in which case an alternative activity had been set up for them. Prior to the trial, participants (senders and receivers) completed a battery of measures. Participants were greeted on arrival and escorted to a reception room that had been specially prepared with comfortable chairs, a coffee table, rugs and curtains so as to make participants feel as comfortable and relaxed as possible prior to the trial. Experimenters encouraged an informal and positive atmosphere, discussing the procedure and answering any questions arising while sharing refreshments. Participants were not informed of the REG task. Participants were then given a guided tour of the facility as the roles of sender and receiver were again explained.

With the assistance of the sender, the experimenter prepared the receiver for the ganzfeld and wished them success. The receiver was seated in a reclining chair and encouraged to relax. They were invited to remove their shoes and cover themselves with a blanket if desired. The receiver wore headphones that had a microphone attached through which they could communicate with the experimenter and be heard by the sender. Halved ping-pong balls were placed over their eyes and held secure with micropore tape. A red light was shone on the receiver's face, positioned immediately in front of them at a distance that was comfortable for them (typically one metre). A computer program was initiated that would sample an REG during the sending period. The receiver was then locked in the room and the sender was guided back to their room.

The trial commenced with the receiver listening to and following a series of progressive relaxation instructions. At the end of the relaxation period the computer program determined whether the session would be a sender or no sender trial. The receiver and experimenter heard a pre-recorded message indicating whether the trial would involve a sender or no sender condition[3]. For no sender trials the sender received an on-screen message asking them to remove their headphones and move over to the other computer in the sender's room to complete an alternative task. The monitor in the sender's room did not show the target clip during no sender trials. For sender trials the sender watched a randomly selected video clip that was played fifteen times with one-minute intervals between plays. Drawing materials were provided for the sender should they wish to sketch elements of the target clip during these 'quiet' periods. During this thirty-minute mentation period the receiver listened to white noise being played through their headphones and reported on any impressions or sensations that they experienced. The experimenter listened to the receiver's mentation via headphones from the experimenter's room and took notes. In the sender condition, the sender could also hear any comments made by the receiver during the mentation period. The REG was sampled once per second for the duration of the sending period, generating a number between 1 and 768, corresponding to one of the clip descriptions. Subsequently, the 20 statements that were most often selected were automatically combined to produce a text file that constituted the REG mentation. In the event of ties, the statement that had been selected first was included.

Following the mentation period, the experimenter reviewed the receiver's mentation. Simultaneously, the sender completed a questionnaire concerning their interaction with the target and sending strategies employed. At the judging stage the receiver was asked to remove their eye-shields but was encouraged to remain in a relaxed state as they watched four video clips, giving each one a percentage similarity rating. After viewing all four clips, they were able to view any or all of them as many times as they wished and to alter their ratings if necessary. The sender was able to listen to the clip soundtracks and the interaction between the receiver and experimenter during the judging stage, but did not view the dummy clips. Once the receiver was satisfied with their ratings, these were confirmed and saved as a permanent record. Only after the data were saved was the target clip revealed and replayed. The sender, experimenter and receiver

[3] This message could have been true or false, and was intended to manipulate the receiver's expectancy. This is not a focus of the present study and the interested reader is referred to Roe, Sherwood & Holt (2003).

then convened for a discussion and debriefing session in the receiver's room. A copy of the trial data record was printed off and signed by all parties to confirm the details of the session.

After completion of all trials in the series, two judges who were otherwise uninvolved with the study[4] independently rated the four clips in each set according to the degree to which they reflected the content of the REG mentation (in the same manner as described for the receiver above).

RESULTS

REG data were not available for four sender trials because of technical malfunctions or failure to initiate the program. Analyses are presented for the remaining 36 trials.

Inter-rater reliability in ranking clips

We were concerned to assess the degree to which the independent judges' ratings would be determined by the mentation, with the alternative being that there was sufficient scope for 'interpretation' that judging may reduce to an elaborate forced choice ESP task for the independent judge. Comparing the target rankings of the two independent judges gives a Cohen's kappa[5] of .202, which is considered poor. As a consequence it was decided not to combine the judges' ratings but rather to consider them separately.

Sender/no sender effects upon ganzfeld/REG performance

The ranks allocated to target clips by the receiver are reported in Table 1. These data have already been described in detail by Roe, et al. (2003) and are included here primarily for comparison purposes. Also included in Table 1 is a summary of the ranks allocated to the target by the independent judges using the REG-generated mentations. We can see from this that the overall hit rate achieved by the receivers in the Ganzfeld, at 25%, is exactly as one would expect by chance. Performance on trials with a sender was only marginally better than on those without (26.1% versus 23.5%). For the independent judges, JW — who had worked on our previous study — gave an above MCE overall direct hit rate (30.6%), whereas our newly-recruited judge, RD, gave a below MCE direct hit rate (16.7%). For both independent judges the hit rate for the sender condition was superior to that for the no sender condition but not significantly so (for JW, 42.1% versus 17.6%, $\chi^2 = 2.54$, $p = .4$; for RD, 26.3% versus 5.9%, $\chi^2 = 1.59$, $p = .6$).[6]

In terms of sums of ranks, better Receiver performance was found in no sender trials, although these differences are not significant ($Z = .639$, $p = 261$). For the independent judges there is little difference in their sums of ranks (for JW, $Z = 1.11$, $p = .134$; for RD, $Z = .824$, $p = .206$). Figure 1 illustrates the range of *z*-scores for target ratings awarded by the independent judges. This confirms the nonsignificant trend in favour of sender versus no sender trials.

---

[4] We are grateful to Russell Davey and Jacqui Wilson for serving as independent judges.

[5] An anonymous reviewer queried the use of Cohen's kappa in preference to Spearman's rank correlation to calculate inter-rater agreement. In this decision we followed Clark-Carter's (1997, p. 533) suggestion that kappa is preferable since Spearman's rank correlation simply measures the direction in which two sets of scores move relative to each other rather than actual agreement. Take for example the case in which Rater 1 ranks the target clips for four trials as 1, 2, 1 and 3, whereas Rater 2 ranks them 2, 3, 2 and 4; these scores would exhibit a perfect (nonparametric) correlation despite the raters showing no agreement at all.

[6] We are grateful to an anonymous reviewer for pointing out that if the ratings from both judges are combined then the hit rate for sender trials is significantly higher than that for no sender trials (13 hits and 25 misses versus 4 hits and 30 misses, $\chi^2 = 5.01$, $p = .03$).

Table 1: A comparison of target rank frequencies for the receiver- and REG-based mentations

		Rank							p
		N	1	2	3	4	SOR	Z	(2-tail)
Receiver	Sender	23	6 (26.1%)	3 (13.0%)	5 (21.7%)	9 (39.1%)	63	-.559	.509
	No sender	17	4 (23.5%)	8 (47.1%)	0 (0.0%)	5 (29.4%)	40	.434	.667
Independent Judge JW	Sender	19	8 (42.1%)	2 (10.5%)	5 (26.3%)	4 (21.1%)	43	.821	.412
	No sender	17	3 (17.6%)	2 (11.8%)	8 (47.1%)	4 (23.5%)	47	-.868	.384
Independent Judge RD	Sender	19	5 (26.3%)	5 (26.3%)	7 (36.8%)	2 (10.5%)	44	.616	.535
	No sender	17	1 (5.9%)	8 (47.1%)	3 (17.6%)	5 (29.4%)	46	-.651	.516

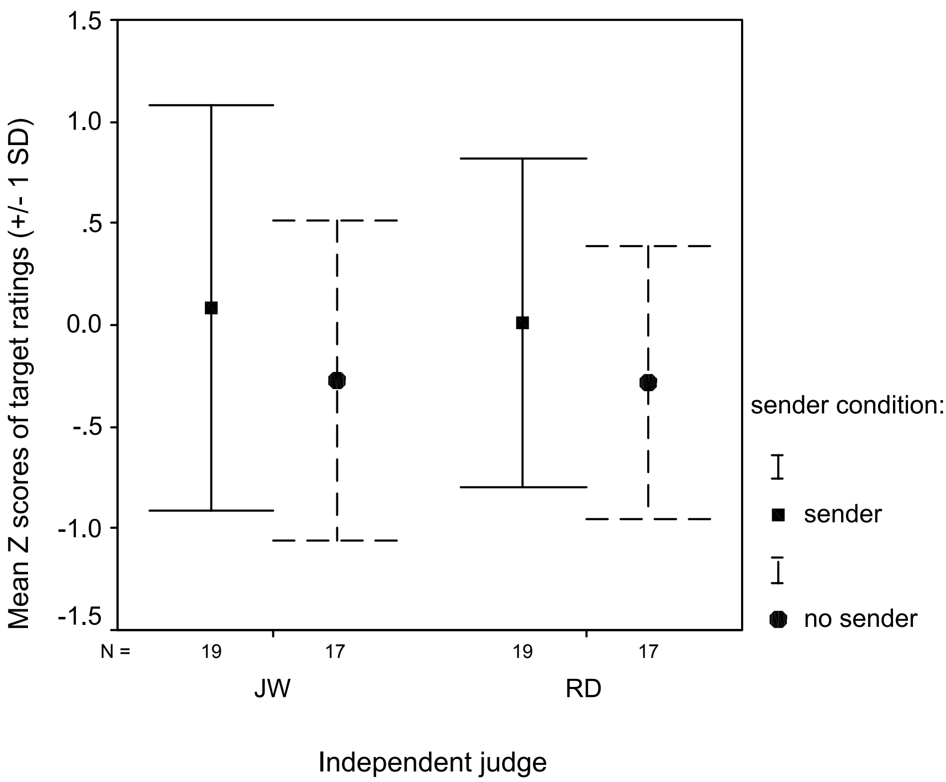


Figure 1: mean Z scores of target ratings for sender and no sender trials

Covariation of REG performance with Receiver performance

Previously we reported a non-significant positive correlation between the independent judge's ratings based on the REG mentation and those by the receiver using their own mentation. Here there was even less agreement between the receiver and independent judges (for JW, κ = -.174; for RD, κ = -.186). This is thus not consistent with the notion of a general sender effect that is reflected in both the receiver's and the REG-based mentation.

Covariation of performance with Sender personality variables

When considering covariation of psi task performance with personality and attitudinal factors it was planned to use the *z*-score of the target clip's similarity rating as the outcome variable. This was deemed preferable to using the simple rank, since it is more sensitive and allows for greater variance across participants, which is essential when considering covariation. The relationships between sender personality and attitude scores and performance for both the receiver and the REG are summarised in Table 2. We can see that only one of the correlations with receiver ratings exceeds .25, which represents a similar pattern to that found previously (Roe, Holt & Simmonds, 2003). Collectively they offer little promise of identifying actual correlates of performance.

Table 2: Spearman rho correlations between Sender personality and attitude measures and performance (and 2-tailed probabilities)

	Receiver ratings z scores	REG-based ratings z scores (JW)	REG-based ratings z scores (RD)
Do you think you can be psychic in the conditions of this study?	.167 (.330)	-.006 (.972)	.037 (.839)
Sender's score on belief in paranormal scale	.259 (.128)	.112 (.534)	.176 (.327)
How creative are you?	-.012 (.942)	.001 (.994)	-.117 (.508)
EPQ neuroticism	-.053 (.764)	.210 (.249)	.112 (.541)
EPQ extraversion	-.070 (.685)	-.006 (.974)	-.093 (.415)

Covariation of performance with Sender strategy variables

Finally, we were interested to see whether the sender's choice of strategy could be used as a predictor of task success. These analyses are presented in Table 3. We can see that once again there are no relationships with receiver performance but that there is a suggestive relationship with REG performance and method of focusing, with better results when using drawing materials to re-enact aspects of the target. There is a significant correlation with method of sending, with episodic sending giving rise to better performance. Although these effect sizes are relatively healthy (>.4), they are clearly only marginally significant at best and would be reduced to nonsignificance if corrected for multiple analyses. Nevertheless they may prove interesting if reproduced in subsequent replications.

Table 3: Spearman rho correlations between Sender strategy measures and performance (and 2-tailed probabilities)

	Receiver ratings z scores	REG-based ratings z scores (JW)	REG-based ratings z scores (RD)
Sending strategy (active versus passive)	.088 (.711)	-.237 (.344)	-.194 (.441)
Holistic or atomistic (whole clip versus elements)	.221 (.336)	-.354 (.149)	-.158 (.531)
Focus (on the target versus on the receiver)	.053 (.821)	.141 (.578)	.159 (.529)
Aspects of clip sent (actual content versus associations with the receiver)	.281 (.217)	.229 (.361)	.115 (.650)
Sending continuous (continuous versus episodic)	.263 (.249)	.281 (.259)	.471 (.049)
Focus tools (mental images versus images and drawings)	-.153 (.508)	-.036 (.886)	-.406 (.094)

DISCUSSION

In this study receivers were only able to correctly select the target clip on 25% of trials, and so were not able to replicate the previously reported above-chance performance of the receiver (Roe, Holt & Simmonds, 2003). Neither were we able to replicate the suggestive overall finding of that study, in which an independent judge was able to produce a 32.5% hit rate based only on REG-derived mentations, although the independent judge who served previously did give a similar hit rate (30.6%). Our newly-recruited judge, RD, gave a below MCE direct hit rate (16.7%).

However, this tells only part of the story, since in this study half the trials did not involve a sender and were considered here to constitute a control condition in which there would be no REG effect and perhaps a lower hit rate for receivers. It was true that the receiver hit rate was slightly higher in the sender condition than in the no sender condition, but this difference is only marginal (26.1% versus 23.5%) and does not provide confirmation of the previous sender superiority (e.g., Raburn & Manning, 1977; Sargent et al., unpub, cited in Milton, 1988-9). For independent judges there seems to be an advantage in favour of sender trials, with the hit rate for JW at 42.1% comparing favourably with the previous study (admittedly over only 19 trials and not significantly different from MCE). Although RD scored below chance overall there was again a nonsignificant advantage for sender trials (26.3% hit rate versus 5.9%). In terms of sums of ranks, the associated effect sizes for JW and RD for the sender trials are positive, at .188 and .141 respectively, but smaller than reported previously (.234). Nevertheless, on the basis of results for sender trials it might be argued that the findings are sufficient to warrant further study, and two more experiments are planned.

Study power is a perennial issue with a protocol as labour intensive as that adopted here. Clearly in ideal circumstances it would be desirable to recruit larger samples and run more sessions. However, it has proven difficult to co-ordinate and run a sufficient number of sessions for studies to be completed in a reasonable time period (the current study was completed in 12 months). It may be worthwhile to consider whether pairs of participants are needed with this procedure. If senders were accurately briefed that their task was to influence a computer generated mentation then this would obviate the need for a receiver participant and for the Ganzfeld induction that goes along with their role. This would give rise to a simplified protocol for which participants could be recruited and scheduled singly. It also offers the opportunity to give immediate feedback to participants on performance by presenting them with the selected statements, which may outweigh any increase in scepticism on the part of participants in having to interact with a machine rather than another human.

The analysis of sender personality, attitude and experiential variables were again disappointing and suggest that the measures used to date are not adequately tapping into individual factors that bear on success at the task. It may be that other measures will prove more successful. At this stage it seems equally likely that pencil-and-paper measures are inadequate *per se* to gauge the complex interplay between factors.

ACKNOWLEDGEMENT

We would like to gratefully acknowledge the financial support of the Institut für Grenzgebiete der Psychologie und Psychohygiene e.V., which has enabled us to conduct this study and to the Bial Foundation for support during analysis and writing up.

REFERENCES

Bem, D.J., & Honorton, C. (1994). Does psi exist? Evidence for an anomalous process of information transfer. *Psychological Bulletin*, 115, 4-18.

Braud, W., & Schlitz, M. (1991). Conscious interactions with remote biological systems: Anomalous intentionality effects. *Subtle Energies*, 2, 1-46.

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove, East Sussex: Psychology Press.

Dunne, B. J., Warnock, E., & Bisaha, J. P. (1977). Ganzfeld techniques with independent ratings for measuring GESP and precognition. In J. D. Morris, W. G. Roll & R. L. Morris (Eds.) *Research in Parapsychology 1976* (pp. 41-43). Metuchen, NJ: Scarecrow Press.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences*, 6, 21-29.

Honorton, C. (1985). The Ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology*, 49, 3-49.

Honorton, C. (1995). Impact of the sender in ganzfeld communication: Meta-analysis and power estimates. *Parapsychological Association 38th Annual Convention Proceedings of Presented Papers*, pp 132-140.

Kanthamani, H., & Palmer, J. (1993). A ganzfeld experiment with 'subliminal sending'. *Journal of Parapsychology*, 57, 241-257.

Milton, J. (1988-9). A possible 'directive' role of the agent in the ganzfeld. *European Journal of Parapsychology*, 7, 193-214.

Morris. R. L., Dalton, K., Delanoy, D. L., & Watt, C. (1995). Comparison of the sender / no sender condition in the ganzfeld. *The Parapsychological Association 38th Annual Convention Proceedings of Presented Papers*, pp. 244-259.

Raburn, L., & Manning, R. (1977). Sender relaxation and expectation in telepathy. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds.) *Research in Parapsychology 1976*. Metuchen, NJ: Scarecrow Press.

Radin, D. I., & Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Foundations of Physics*, 19, 1499-1514.

Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., & Greenwood, J. A. (1966). *Extrasensory perception after sixty years.* Boston: Bruce Humphreys. (Originally published in 1940).

Roe, C. A. (1996). Clients influence in the selection of elements of a psychic reading. *Journal of Parapsychology*, 60, 43-70.

Roe, C. A. (1998). Belief in the paranormal and attendance at psychic readings. *Journal of the American Society for Psychical Research*, 90, 25-51.

Roe, C. A., Holt, N., & Simmonds, C. A. (2003). Considering the sender as a PK agent in ganzfeld ESP studies. *Journal of Parapsychology*, 67, 129-145.

Roe, C. A., & Sherwood, S. J., & Holt, N. J. (2003). Interpersonal psi: Exploring the role of the sender in Ganzfeld GESP tasks. *Proceedings of the Parapsychological Association 46th Annual Convention*, 237-256.

Steinkamp, F., Boller, E., & Bösch, H. (2002). Experiments examining the possibility of human intention interacting with random number generators: A preliminary meta-analysis. In C. Watt (Ed.), *Proceedings of presented papers: The Parapsychological Association 45th Annual Convention*, 256–272.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring the Sheep-Goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research*, 59, 172-186.

Utts, J. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration*, 10, 3-30.

Williams, C., Roe, C. A., Upchurch, I., & Lawrence, T. R. (1994). Senders and geomagnetism in the autoganzfeld. *The Parapsychological Association 37th Annual Convention Proceedings of presented papers*, 429-438.

Address for correspondence: Chris A. Roe,
Division of Psychology, University College
Northampton, Park Campus, Boughton Green Road,
Northampton NN2 7AL, UK.
E-mail: chris.roe @northampton.ac.uk

FIELD STUDY OF AN ENHANCEMENT EFFECT ON LETTUCE SEEDS: WORKING IN ADVERSE CONDITIONS

S.M. Roney-Dougal[1] & J. Solfvin[2]
[1]*Psi Research Centre, Glastonbury, Britain*
[2]*Center for Indic Studies, University of Massachusetts, USA*

ABSTRACT

During the years 2000 – 2002, research on organic farms looked at the effect of a healer on lettuce seeds. The basic hypotheses were that the healer would enhance the seeds to produce greater yield and greater health. The first year (Roney-Dougal & Solfvin 2002a) found a significant result for the second hypothesis, as measured by looking at the fungal damage. The second year (Roney-Dougal & Solfvin, 2002b, 2003) found significant results on all the measures, greater yield, less fungal damage and less slug damage.

However, doing field trials has many problems compared with doing laboratory research, weather being one of the major variables that cannot be controlled. Organic farmers are working under conditions of immense stress, working very long hours, especially during the summer season.

For various reasons, in 2001 trials 6 and 7 were not planted out until long after they had become pot-bound. The delay in planting out was so great for trial 6 that the plants never grew properly, and trial 7 was harvested first as those plants grew to a point where they became big enough for sale. In the end trial 6 were harvested after they had been frosted just so that some data could be collected, but with no benefit to the farm.

In 2002 the research took place on another farm and was a disaster. For various reasons up to half of the lettuce plants died in the seedling trays, and the people working on the farm became "spooked." They considered that the reason for the loss of plants was due to the research trials upsetting the delicate balance of energy needed to keep the farm running smoothly. Therefore the research was terminated after only two trials were harvested.

However, in both 2001 and 2002 significant results were still found. Trials 6 & 7 in 2001 both showed first place "hits" for gross weight of lettuces in the "enhanced" (HX) condition, which, when combined with trials 1 – 5, give an average rank for gross weight of 1.92, which yields $t(11) = -2.24$ and two-tailed $p = .046$. The net weight was also significant, with rank = 1.92, $t(11) = -2.55$ and two-tailed $p = .027$. When each harvest is ranked, by group, on the number of lettuces produced, the HX group has mean rank of 1.79, with $t(11) = -2.75$, and two-tailed $p = .019$. The total yield was also significant for the HX group, which had average rank of 1.83, and $t(11) = -2.77$, for two-tailed $p = .018$. The two trials in 2002 yielded nine experimental vs. control pairings. The nine "experiments" comparing net weight HX to another condition yield an average effect size (r) of .21 (sd = .31), which is statistically significantly greater than chance expectation ($t(8) = 1.97$, $p = .042$, one-tailed). The six control experiments (NH v. controls) yield average effect size $r = -.02$ (sd = .26), which is not significantly different from zero ($t(5) = -0.20$, $p = .43$, one-tailed).

INTRODUCTION

Some of the first psychic healing experiments in parapsychology involved a healer attempting to influence plant germination and growth. For a review of healing studies see Benor (1993, 2004). The basic laboratory design from these experiments has been used for large scale field trials on organic farms in England. These have been run from 2000 – 2002, the first two years at Radford Mill farm in north Somerset, and in 2002 at South Farm (pseudonym) in south Somerset. On organic farms no fungicide, pesticides or artificial fertilisers are used. This can mean that farmers have problems with pests, with fungal diseases and a lower overall yield. The basic aim of these trials was to see whether a laboratory experiment in which a healer attempted to "enhance" seed germination, health and growth of the resulting plants, would translate into a practical application for farmers by reducing pest and disease problems and enhancing yield. At Radford Mill each trial was harvested in two separate batches one week apart, called the first and second harvests which were analysed separately. In 2000 we found significantly less fungal growth ($F(3,24) = 3.13$, $p = .044$) on the plants; in 2001 average gross weight for first harvests is larger than chance expectation (ES =

.09), though not significant (Z=1.47, p = .072), but is significant for second harvests (ES = .19, Z=3.04, p < .001). Net weight for the enhanced group is significantly greater than chance for both first (ES = .16, Z=2.49, p < .01) and second harvests (ES = .15, Z = 2.28, p = .011). Average slug damage ratings are reduced significantly for first harvests (ES = -.11, Z = -1.66, p < .05), and for second harvests (ES = -.12, Z = -1.82, p < .05). Average fungal damage is statistically significantly reduced for first harvests (ES = -.13, Z = -2.11, p < .05), but not significantly so for second harvests (ES = -.03, Z = -0.41, p = n..s.). Overall, the effect sizes are in the "small" range, from .09 to .19 across the five plantings. However, the strength of these results is augmented by a remarkable level of consistency in directionality. This is precisely the sort of consistency which can, over time, cumulate into a substantial amount, and suggests that a healer can enhance seeds sufficiently to make a commercial difference to the farm.

However, organic farms are not laboratories and conditions cannot be specified. The research has to make allowance for the needs and problems encountered in farm conditions. In 2001 two trials (trials 6 and 7) were not included in the data analysis prior to seeing the data, because the plants were not planted out in the field in time for a normal harvest to take place before the winter. These plants were however harvested by the farm so that the data could be collected, even though the plants from trial 6 were not any use for them as they were frosted before harvesting. This meant that the slug and fungal damage could not be assessed properly as the lettuces were in such bad condition. However the data was analysed almost one year later to see whether or not the effect of the healer held up under these adverse conditions. The method and procedure for this data is identical with that in the previous two papers and so will only briefly be described here. For full details please see Roney-Dougal & Solfvin (2002a) and Roney-Dougal & Solfvin (2003).

In 2002 trials were run on a new farm, as the original farm wished to take a break from field trials, which cause a considerable amount of extra work. This new farm suffered a considerable loss of plants and so asked for the trials to be terminated. The design and methodology with these trials varied considerably from the previous ones because the farm grew far fewer lettuces in total, and grew a large number of different varieties. In the previous year it had been noticed post hoc that the data from trials 2 and 3 were more significant than the other trials. In these two trials there had been a different variety of lettuces used, because the grower had suggested we use this other variety, as it would show up the effects of fungal damage more clearly than the variety we had been using. As this new farm grew many varieties, it seemed an opportunity for us to look at the effects of a healer on different varieties as a controlled experiment. JS therefore designed a programme in which each of the five varieties grown would be an experimental trial, or one of the two different types of control trial, counterbalanced over 10 trials.

This paper is in two parts: the final two trials from 2001 at Radford Mill Farm, and the two trials harvested in 2002 at South Farm.

RADFORD MILL – METHOD

Hypotheses

As with the previous two years there were two primary hypotheses:

1) The "enhanced" seeds will have greater growth than the control.

2) The "enhanced" seeds will have better health than the control.

3) In addition we looked to see if there would be differential effect on the different varieties. There was no specific hypotheses, this being exploratory, but a suggestion that different varieties would respond differently to the effect of the healer dependant on their resistance to pests and disease.

The following four outcome variables were pre-planned:

1) Growth variables: gross and net weight of the plants were measured when harvested;

2) Health variables: the plants were checked for slug and for fungal damage, and rated on a five-point scale.

Enhancement Procedure

For each trial, in the packing shed, in the presence of the experimenter (SRD) a person, who acted as the randomiser, was given a sealed pot of 1000 red oak lettuce seeds (Valdai). They counted 100 seeds, into each of four identical jars, making a total of 400 experimental seeds in each trial. These were then closed with a screw top lid. The randomiser was a person who knew no one who worked on the farm, and had no contact with the farm other than on the one occasion they performed the function of ensuring that the four jars of seeds were randomly ordered. At this point nobody knew which jar would be given to the healer – all four jars were identical with no identification. One jar of seeds (HX) was given by SRD to the healer (MP). One jar (NH) was given by SRD to the control person, who handled it exactly as MP handled his. This person claims no healing ability and mimicked the actions of MP whilst he was doing the enhancement. This person also performed this function the previous year as she lives on the farm, has no involvement in the vegetable side of the farm and was very interested in the research. Two jars (C1 & C2) were untreated controls and were taken by SRD and placed outside the door of the shed, where she stood whilst the healer performed the enhancement to make sure that there was no disturbance. The rationale behind this was that Schwartz (1990) found that controls in the same room as the healer can be affected. This was not taken into account in the preliminary experiment, and was included as a variable here, as both SRD and the healer felt it might be valid. However, as these trials were not included with the 5 previous trials no comparison can be made. This methodology suffers from a possible decision augmentation theory (DAT) problem (May, Utts & Spottiswoode, 1995). Essentially DAT suggests that perhaps the psi happened by the randomiser putting all the best seeds in one jar and then SRD gave this jar to the healer. The methodology was amended in the design for 2002 so that this could not occur.

MP enhanced the seeds by holding his hands approximately one foot from the jar. Each enhancement lasted about five minutes. MP is a professional healer working for the National Federation of Spiritual Healers, the most respected organisation of healers in Britain. SRD asked him to help with this research, and he worked with us during all of 2001 and 2002. After the enhancement, when everyone had left the shed, the randomiser assigned labels with the trial number and A, B, C or D to the four jars of seeds, which had been left on the table in the places where each person had sat. As the enhancement sessions for trials 6 and 7 were held one after the other on the same morning, there was the same randomiser for both the trials, and the randomisation was done after both enhancements were complete. (For full details of the randomisation procedure, please see Roney-Dougal & Solfvin, 2002a.) Once the randomiser had labelled the jars they left them in the middle of the table for the grower to collect, left the farm and had no further contact with the farm.

Growing Procedure

The seeds were all sown by one person to ensure uniformity of sowing procedure. Trial 6 was sown on July 3rd and trial 7 on July 10th. The seeds were germinated in a dark shed. The initial growing conditions were in trays in a polytunnel, with the approximately three week old plants being placed in their trays outside where they were left until September when they were finally planted out, the final harvest being in December. SRD recorded the numbers of seeds that had germinated, the number of plants in each row. and the number that were harvested.

Harvesting Procedure

At a time determined by the grower, first of all trial 7 plants were harvested. Because these had been sown two weeks after trial 6, they had been less stressed by being left in the trays for so long, and so grew to a usable size. Trial 6 was harvested after the first frost when they could no longer be left in the field. Each lettuce was harvested by being cut at ground level. These were brought into the packing shed, and a team of five people then assessed the lettuces for health and weight. First the lettuce was assessed for fungal damage, and then for slug damage. This was recorded on a five-point scale, with 1 indicating little damage up to 5 indicating much damage. Then the lettuce was weighed to give gross weight. The outer leaves were trimmed and then they were re-weighed to obtain net weight, which was a lettuce ready for sale.

At the beginning of each harvesting session, the harvesters all independently rated a lettuce for slug and fungal damage, and then checked to see how closely their assessments matched. If there was a difference they discussed this and rated another lettuce, until agreement was reached.

After all the data had been recorded and sent to JS for analysis, the randomiser revealed the codes.

RESULTS

Table 1 summarises the results from Trials 6 and 7, showing the n, mean (Mn), standard deviation (sd), and rank for each measure: gross weight (GW), net weight (NW), slug damage (SL), fungal damage (FG), and Total Yield, by treatment condition.

Table 1: Means, SDs, and Ranks of Measures By Condition.

		GROUP			
		HX	NH	C1	C2
GROSS WEIGHT (GW)					
Trial 6	n	82	81	62	78
	Mn(sd)	73.17 (17.24)	63.09 (19.44)	61.05 (14.40)	69.85 (19.94)
	Rank	1	3	4	2
Trial 7	n	94	90	95	92
	Mn(sd)	93.94 (22.46)	83.50 (19.64)	88.79 (19.78)	90.92 (19.04)
	Rank	1	4	3	2
NET WEIGHT (NW)					
Trial 6	n	82	81	62	78
	Mn(sd)	36.95 (23.02)	31.98 (21.63)	31.77 (18.49)	38.90 (24.74)
	Rank	2	3	4	1
Trial 7	n	94	90	95	92
	Mn(sd)	72.07 (20.15)	63.00 (18.94)	65.00 (20.59)	67.99 (17.62)
	Rank	1	4	3	2
SLUG DAMAGE (SL)					
Trial 6	n	82	81	62	78
	Mn(sd)	0.75 (1.18)	0.57 (0.88)	0.40 (0.68)	0.66 (0.84)
	Rank	4	2	1	3
Trial 7	n	94	90	95	92
	Mn(sd)	1.62 (0.64)	1.12 (0.71)	1.38 (0.57)	1.32 (0.68)
	Rank	4	1	3	2
FUNGUS DAMAGE (FG)					
Trial 6	n	82	81	62	78

	Mn(sd)	3.26 (1.07)	3.18 (1.02)	3.35 (0.98)	2.96 (1.11)
	Rank	3	2	4	1
Trial 7	n	94	90	95	92
	Mn(sd)	2.16 (0.67)	2.18 (0.76)	2.33 (0.72)	2.35 (0.67)
	Rank	1	2	3	4
TOTAL YIELD					
Trial 6	n	82	81	62	78
	Kg	3.030	2.590	1.970	3.034
	Rank	2	3	4	1
Trial 7	n	94	90	95	92
	Kg	6.775	5.670	6.175	6.255
	Rank	1	4	3	2

HX is the "enhanced" group, NH the control mimic group, and C1 &C2 the untouched controls.

These two trials have insufficient data to be analysed using the original rank analysis method (Roney-Dougal & Solfvin, 2002b, 2003). But it is seen that the earlier trend continues – Trials 6 & 7 both showed first place "hits" for gross weight of lettuces in the "enhanced" (HX) condition, and first and second place "hits" for net weight. Adding these two trials to the first five trials reported in Roney-Dougal & Solfvin, (2002b, 2003) we found that the HX group was still statistically significantly more productive than the other groups, with average rank (among groups) for gross weight of 1.92 (where an average rank of 2.5 is expected by chance), which yields $t(11) = -2.24$ and two-tailed $p = .046$. The net weight was also statistically significant, with Mn rank = 1.92, $t(11) = -2.55$ and two-tailed $p = .027$.

The number of lettuces produced was also significantly increased by the healing treatment. When each harvest is ranked, by group, on the number of lettuces produced, the HX group has mean rank of 1.79, with $t(11) = -2.75$, and two-tailed $p = .019$. The total yield was also significant for the HX group, which had average rank of 1.83, and $t(11) = -2.77$, for two-tailed $p = .018$.

Slug damage showed little between group variation and the HX group was unremarkable, just slightly larger (against the hypothesis) than chance expectation (of 2.5) with average rank of 2.58, $t(11) =0.28$, two-tailed $p = .782$. Fungal damage was in the correct direction, with average rank of 2.13, but not significantly so ($t(11) = -1.15$, two-tailed $p = .28$).

Thus, we can conclude from this augmented data that there was indeed a clear effect of the healer's treatment on the growth of the lettuces, although not on the health parameters (resistance to slug and fungal damage) included in this research. It is particularly interesting that the positive effect on the lettuce growth endured the on-coming cold weather, being roughly the same in trials 6 & 7 as in the earlier, height-of-the-growing-season trials. Although the lettuces produced in these final two plantings were overall substantially smaller than in earlier plantings, the ones in the HX group outgrew the other groups as was the case throughout the season.

DISCUSSION

Thus, despite the adverse conditions, we still see the positive effect of enhancement on the plants in these two trials. The lack of fungal damage results for trial 6 is actually an artifact. Owing to the frost damage which the plants incurred, the assessment of fungal damage was virtually impossible, the harvesters saying that they could not distinguish whether the damage to the leaves was frost damage or fungal damage.

Also because of the cold there was virtually no slug damage as slugs are not active at this time of year, and so a proper assessment could not be made for this parameter.

Therefore this shows that a healer can be of assistance to a farm under normal and under adverse conditions.

South Farm – Method

Enhancement Procedure

JS created a design in which only he knew which plants were from the enhanced seed, whereas in the previous design the randomiser had held that information. His new design also ensured that there could be no effect of intuitive data sorting (DAT). In this design there were five varieties of lettuce and each variety was divided into two parts, A and B. Therefore there are ten sets of seeds, Var.1A, Var. !B, Var. 2 A, Var. 2B, etc., each in their own plastic bag. There were 72 seeds in each set making a total of 720 seeds in each trial.

JS gave the person, who acted as the "randomiser," codes for each trial so that each set could be placed at random into a Jiffy bag, labeled from A to J. JS also gave codes to SRD which specified which Jiffy bags would be enhanced by the healer (MP), which would receive the mimic treatment by the control person and which would be put aside with no treatment of any kind. Thus the randomiser did not know which Jiffy bags were treated and SRD, MP, and the mimic did not know which varieties were in each bag.

The enhancement procedure differed slightly from that used at Radford Mill. First, SRD went in to the packing shed to ensure that the table was clear and clean, that there was space for MP to stand at one end of the table and the mimic to stand at the other, and that the seeds were in their plastic bags in order. She left the shed and informed the randomiser that all was ready. Next, the randomiser entered the packing shed and placed the seeds in their appropriate Jiffy bags, according to the instructions given by JS, which were in a sealed envelope. These were left on the table in the packing shed where the enhancements took place. SRD, MP, and the control person then entered the shed. SRD opened her envelope with instructions from JS and gave MP and the control their respective three Jiffy bags each, removing the remaining four from the table. MP did his enhancement, giving each of the three Jiffy bags in each trial its own treatment, apart from the first trial when he "enhanced" all three simultaneously. He spent between two to five minutes on each bag, most bags receiving approximately three minutes, SRD timing this. The mimic focused on copying his actions, which were the same as at Radford Mill, holding his hands approximately twelve inches from the bag and angling the palm towards the bag. The control person was recruited by SRD, and was a different person for each trial. They claim no healing ability and no particular ability with plants. When all enhancements were complete the Jiffy bags were again placed in order on the table and all three people left the packing shed. The randomiser then entered, removed the plastic seed bags from the Jiffy bags and left them on the table for the grower to collect.

Growing Procedure

The seeds were all sown by one person to ensure uniformity of sowing procedure. The seeds were germinated at one end of the greenhouse. Unfortunately the person responsible for this had not worked on a farm before and was very new at the job, having worked previously in an office. She did not realise that all the seeds needed equal treatment and so some of the trays were placed underneath other trays resulting in some of the germinating seeds "bolting," which means that they grew very tall and spindly during this phase. This resulted in differing germination conditions for different sets.

The initial growing conditions were in trays in the same greenhouse, either on a shelf or on the ground, with the approximately three week old plants being placed in their trays in a different greenhouse or outside, where they were left to harden off until they were planted out. There were problems with this phase also and this is the phase when most plants died.

A photographer who had worked with us the previous two years at Radford Mill again kept a photographic record of the seedlings in the greenhouse, and the plants growing in the fields.

The plants were planted out approximately six weeks after sowing in two rows, with varieties 1A - 5 A on one side, and varieties 1B - 5 B on the other side. Thus all plants had equal growing conditions because all plants were outside on one side and next to another lettuce plant on the other, apart from those four plants at the end of the row. Again there were problems here. Because of the unequal numbers of plants reaching this stage, there tended to be more of either A or B and so the variety last to be planted was often in a single row on its own, which gave unequal growing conditions. The first two trials were planted out simultaneously and, because of the very cold weather, grew very slowly so that they were badly affected by weeds.

Harvesting Procedure

At a time determined by the grower, the plants were harvested from the variety that was ready, attempting to harvest equal numbers of A and B within that variety at each time. The plants were weighed, assessed for slug and fungal damage and trimmed for packing and re-weighed as in previous years, with people who had done this procedure before.

Because of market requirements instead of one harvest each week, there were three harvest a week with smaller numbers being harvested each time, these numbers varying according to the orders received by the farm. As there were only a few lettuces harvested on each occasion only one person did the assessments on any one trial. In all there were three harvesters (SE, SH and MA).

RESULTS

1: The plant growth measures (weights).

In sorting out the comparison groups (A, B), the four treatment conditions (HX, NH, C1, C2), the five varieties (1-5), the three different harvesters (SH, MA, SE), and the various dates of harvesting, there were a few essential principles needed to do so. First, we had to establish the basic unit of the analysis. For this data, it is an A v. B paired comparison, for which the variety, harvester, and date of harvesting are constant. There were six (6) such comparisons in trial #1, and nine (9) in trial#2, and that's the basic data we have to work with, shown in tables 2 and 3. A few of the groupings provided could not be used because they included crops harvested on different days, or by different harvesters. We also lost some crop data – but very little – because a few plants here or there were harvested without an appropriate pairing.

Each one of these A v. B comparisons is itself a small experiment, with an associated t-test and p-value (in the tables). Table 2 shows two such "experiments" for Variety 1, and one experiment each for the other four varieties. But do note that only the first four "experiments" in Table 2 involve HX v. another group, thus the last two experiments in Table 2 are controls. Two of the four HX comparisons significantly support the hypothesis, while neither of two control comparisons (NH v. control) are in the correct direction, and are not statistically significant.

Table 2: Trial 1 – Mean (SD) Crop Weights By Condition (Hx, NH, C1, C2) By Variety (1 – 5)

VARIETY (Date/harvester)	Wt.	HX	NH	C1	C2	t-test
Variety 1 (25.7.by MA)	Gross	278.2 (82.1)	--	269.1 (81.9)	--	t(54) = 0.42
	Net	223.5 (82.5)	--	215.7 (78.3)	--	t(54) = 0.36

	n	28	–	28	–	
Variety 1 (26.7.by SE)	Gross	226.0 (78.5)	–	174.9 (65.7)	–	t(46) = 2.45*
	Net	203.2 (72.8)	–	150.7 (63.3)	–	t(46) = 2.67**
	n	23	–	25	–	
Variety 2 (19.7. by SH)	Gross	300.0 (111.6)	–	–	354.9 (147.1)	t(19) = -0.91
	Net	275.1 (86.9)	–	–	313.1 (131.6)	t(18) = -0.68
	n	8[1]	–	–	13	
Variety 3 (22.7. by SH)	Gross	367.7 (99.6)	286.0 (94.0)			t(31) = 2.20*
	Net	326.6 (91.7)	252.6 (84.6)			t(31) = 2.18*
	n	23	10			
Variety 4 (22.7. by SH)	Gross	–	275.0 (108.1)	306.5 (98.2)	–	t(50) = -0.95
	Net	–	244.2 (103.0)	278.2 (96.7)	–	t(50) = -1.06
	n	–	12	40	–	
Variety 5 (15.7. by SH)	Gross	–	168.5 (68.5)	–	175.2 (66.5)	t(40) = -0.32
	Net	–	144.7 (68.3)	–	146.1 (62.2)	t(40) = -0.07
	n	–	17	–	25	

[1]N = 8 for gross weight only – one missing value for Net weight.
p = .05; ** p = .01
HX is the "enhanced" group, NH the mimic control group, and C1 &C2 the untouched controls.
MA, SE & SH are the harvesters

Table 3 shows the second trial to be even stronger – four of the five HX comparisons are in the predicted direction and three of them highly statistically significant. The one aberrant comparison (variety 3, harvested by MA on 29.7), is in the wrong direction and statistically significant. Nonetheless, the HX comparison group looks good compared to the four control experiments, of which only one is in the correct direction (NH > C1), and is also the only statistically significant one.

Table 3: Trial 2 – Mean (SD) Crop Weight By Condition (Hx, NH, C1, C2) By Variety (1 – 5)

VARIETY (Date/harvester)	Wt.	HX	NH	C1	C2	t-test
Variety 1 (26.7. by SE)	Gross	197.7 (71.1)	–	102.0 (56.1)	–	t(13) = 2.92*
	Net	175.7 (68.2)	–	89.8 (51.9)	–	t(13) = 2.77*
	n	7	–	8	–	
Variety 1 (29.7. by MA)	Gross	148.2 (55.3)	–	98.7 (61.1)	–	t(92) = 4.11***
	Net	111.1 (49.8)	–	70.8 (42.1)	–	t(92) = 4.18***
	n	52	–	42	–	
Variety 2 (15.7. by SH)	Gross	289.2 (57.6)	–	–	278.8 (44.4)	t(70) = 0.86
	Net	233.1 (45.2)	–	–	222.3 (37.2)	t(70) = 1.10
	n	36	–	–	36	
Variety 2 (18.7. by MA)	Gross	421.4 (95.0)	–	–	308.4 (77.3)	t(53) = 4.85***
	Net	296.5 (65.3)	–	–	234.9 (64.3)	t(53) = 3.53***
	n	27	–	–	28	
Variety 3 (29.7. by MA)	Gross	174.9 (76.1)	234.2 (114.4)	–	–	t(62) = -2.39*
	Net	129.1 (62.2)	193.5 (105.8)	–	–	t(62) = -2.97**
	n	47	17	–	–	
Variety 4 (1.8. by SE)	Gross	–	429.9 (128.8)	431.9 (101.6)	–	t(54) = -0.07
	Net	–	366.1 (112.6)	350.5 (90.6)	–	t(53) = 0.57
	n	–	28[1]	28	–	
Variety 4 (2.8. by SE)	Gross	–	488.9 (143.0)	377.9 (84.5)	–	t(32) = 2.76**
	Net	–	424.1 (129.1)	307.3 (92.4)	–	t(32) = 3.03**
	n	–	17	17	–	
Variety 4 (5.8. by SH)	Gross	–	575.9 (122.4)	601.1 (51.9)	–	t(25) = -0.64

	Net	–	491.0 (120.9)	543.1 (54.8)	–	t(25) = -1.33
	n	–	16	11	–	
Variety 5 (5.8. by SH)	Gross		401.9 (117.5)		415.7 (103.8)	t(67) = -0.51
	Net		350.1 (111.8)		371.9 (106.6)	t(67) = -0.82
			39		30	

[1]N = 28 for gross weight only – one missing value for Net weight.
* p = .05; ** p = .01; *** p = .001
Note. T-values with positive sign are in direction of hypothesis
HX is the "enhanced" group, NH the mimic control group, and C1 &C2 the untouched controls.
MA, SE & SH are the harvesters

In order to summarize these it's easier to use ONLY the net weight parameter, since gross and net weight are strongly correlated. Net weight and slug/fungal damage are correlated because they determine how much has to be cut off. We utilize "effect size" (r.) instead of t-test significance levels. The simplest summary is the overall test of the question, "Did the healer's enhancement treatment work?" The simple answer is YES: The nine "experiments" comparing HX to another condition yield an average effect size (r) of .21 (sd = .31), which is statistically significantly greater than chance expectation (t(8) = 1.97, p = .042, one-tailed). The six control experiments (NH v. controls) yield average effect size r = -.02 (sd = .26), which is not significantly different from zero (t(5) = -0.20, p = .43, one-tailed).

In some ways, the most interesting questions are about the portion of the overall effects that can be attributed to the lettuce variety or the person who harvested it. To survey this, we lay out the effect sizes of the nine HX comparison trials into a three (variety) by three (harvester) array, as shown in Table 4.

Table 4: Effect Sizes by Lettuce Variety and Harvester

Variety	SH		MA		SE		**AVE.**
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2	
VAR 1	–	–	0.05	0.40	0.37	0.61	**0.36**
VAR 2	-0.16	0.13	–	0.44	–	–	**0.14**
VAR 3	0.37	–	–	-0.35	–	–	**0.01**
AVE.	**0.12**		**0.14**		**0.49**		

Had we continued data collection as planned through the season, all of the varieties would have been represented, but since we only have what we have, *no inferences or conclusions can be drawn from the data.* At best, these may suggest hypotheses for future study. No statistical analyses are valid with these data so none are provided.

The analysis of the slug and fungus damage data is far more complicated, as can be seen from tables 5, 6 and 7. No obvious pattern emerges.

Table 5: Trial 1 – Mean (SD) Slug Damage By Condition (Hx, NH, C1, C2) By Variety (1 – 5)

VARIETY (Date/harvester)		HX	NH	C1	C2	t-test
Variety 1 (25.7 by MA)	rating	1.05 (0.64)	–	0.57 (0.59)	–	t(54) = -2.93**
	n	28		28	—	
Variety 1 (26.7. by SE)	rating	0.13 (0.22)	–	0.14 (0.27)	–	t(46) = 0.13
	n	23	–	25	–	
Variety 2 (19.7. by SH)	rating	2.31 (0.59)	–	–	2.31 (0.66)	t(19) = -0.02
	n	8	–	–	13	
Variety 3 22.7. by SH)	rating	2.04 (0.42)	2.25 (0.49)	–	–	t(31) = 1.23
	n	23	10	–	–	
Variety 4 (22.7. by SH)	rating	–	1.83 (0.33)	1.70 (0.45)	–	t(50) = -0.95
	n	–	12	40	–	
Variety 5 (15.7. by SH)	rating	–	1.77 (0.83)	–	1.88 (0.78)	t(40) = 0.46
	n	–	17	–	25	

*p = .05; **p = .01; ***p = .001
Note. T-values with positive sign are in direction of hypothesis
HX is the "enhanced" group, NH the mimic control group, and C1 &C2 the untouched controls.
MA, SE & SH are the harvesters

Table 6: Results of Trial 1 – Mean (SD) Fungal Damage By Condition (Hx, NH, C1, C2) By Variety (1 – 5)

VARIETY (Date/harvester)		HX	NH	C1	C2	t-test
Variety 1 (25.7. by MA)	rating	2.71 (1.07)	–	2.04 (1.42)	–	t(54) = -2.02*
	n	28	–	28	–	
Variety 1	rating	0.83 (0.63)	–	2.04 (1.10)	–	t(46) = 4.61***

(26.7. by SE)						
	n	23	~	25	~	
Variety 2 (19.7. by SH)	rating	1.88 (0.69)	~	~	1.81 (0.88)	t(19) = -0.18
	n	8	~	~	13	
Variety 3 (22.7. by SH)	rating	1.67 (0.54)	1.80 (0.54)	~	~	t(31) = 0.62
	n	23	10	~	~	
Variety 4 (22.7. by SH)	rating	~	1.38 (0.48)	1.68 (0.40)	~	t(50) = 2.27
	n	~	12	40	~	
Variety 5 (15.7. by SH)	rating	~	0.94 (0.90)	~	0.80 (0.71)	t(40) = -0.57
	n	~	17	~	25	

*p = .05; **p = .01; ***p = .001

Note. T-values with positive sign are in direction of hypothesis.

HX is the "enhanced" group, NH the mimic control group, and C1 &C2 the untouched controls.

MA, SE & SH are the harvesters

Table 7: Results of Trial 2 – Mean (SD) Slug and Fungal Damage By Condition (Hx, NH, C1, C2) By Variety (1 – 5)

VARIETY (Date/harvester)	Type	HX	NH	C1	C2	t-test
Variety 1 (26.7 by SE)	slug	0.57 (0.45)	~	0.75 (0.27)	~	t(13) = 0.95
	fungus	1.57 (0.45)	~	1.44 (0.86)	~	t(13) = 0.37
	n	7	~	8	~	
Variety 1 (29.7 by MA)	slug	0.66 (0.88)	~	0.14 (0.35)	~	t(92) = -3.59***
	fungus	3.28 (1.34)	~	3.14 (1.66)	~	t(92) = 0.44
	n	52	~	42	~	
Variety 2 (15.7. by SH)	slug	2.15 (0.64)	~	~	2.21 (0.77)	t(70) = 0.33
	fungus	1.58 (0.91)	~	~	1.43 (0.82)	t(70) = -0.75
	n	36	~	~	36	

Variety 2 (18.7. by MA)	slug	2.22 (0.35)	~	~	2.14 (0.54)	t(53) = -0.64
	fungus	2.57 (0.68)	~	~	1.88 (0.99)	t(53) = -3.05**
	n	27	~	~	28	
Variety 3 (29.7. by MA)	slug	1.59 (0.57)	1.82 (0.53)	~	~	t(62) = 1.50
	fungus	1.87 (0.77)	1.53 (1.01)	~	~	t(62) = -1.45
	n	47	17	~	~	
Variety 4 (1.8. by SE)	slug	~	2.55 (0.77)	2.46 (0.45)	~	t(54) = -0.53
	fungus	~	1.43 (0.73)	2.21 (0.90)	~	t(54) = 3.60***
	n	~	28	28	~	
Variety 4 (2.8. by SE)	slug	~	2.50 (0.59)	2.62 (0.45)	~	t(32) = 0.66
	fungus	~	1.50 (0.64)	1.68 (0.53)	~	t(32) = 0.88
	n	~	17	17	~	
Variety 4 (5.8by SH)	slug	~	1.75 (0.32)	1.82 (0.34)	~	t(25) = 0.54
	fungus	~	0.75 (0.41)	0.46 (0.27)	~	t(25) = -2.10*
	n	~	16	11	~	
Variety 5 (5.8 by SH)	slug	~	2.35 (0.55)	~	2.47 (0.45)	t(67) = 0.97
	fungus	~	0.85 (0.46)	~	0.88 (0.58)	t(67) = 0.30
	n	~	39	~	30	

*p = .05; **p = .01; ***p = .001
Note. T-values with positive sign are in direction of hypothesis.
HX is the "enhanced" group, NH the mimic control group, and C1 &C2 the untouched controls.
MA, SE & SH are the harvesters

Using the same approach as described above for the growth (wt.) data, the summed effect sizes for the experiments do not show any significance effects for slug or fungus damage. For the nine experiments comparing HX with another group, the average effect size is r = -0.01 (sd = 0.23), which is not significantly different from zero (t(8) = -0.14, p = .554, one-tailed). Similarly for the control experiments (NH v. controls) the average effect size is .04 (sd = 0.11) which does not differ significantly from zero (t(5) = 0.78, p = .236, one-tailed).

For fungus damage, the nine HX comparisons averaged r = -0.02 (sd = 0.27) which doesn't differ significantly from zero (t(8) = -0.17, p = .567, one-tailed). The control experiments with fungus damage averaged r = 0.08 (sd = 0.30) which is not significantly different from zero (t(5) = 0.64, p = .277, one-tailed).

DISCUSSION

Once again the growth of the plants was significantly increased when the seed had been "enhanced" by the healer, despite the difficulties experienced by the farm and the research team. This corroborates the previous findings (Roney-Dougal & Solfvin, 2002a, 2003).

The more complicated findings with regard to health of the plants as measured by slug and fungus damage is to be understood in the light of working with seven different varieties. Some of the varieties in trial 1 are not the same as in trial 2, and the different varieties have different properties with regard to susceptibility to fungal damage. Had we been able to continue the harvests for all seven trials then we would have been able to look at the effect of the healer with regard to these breeding traits as originally planned. There are three varieties which are the same in both trials = vars. 1, 3 & 4 (Frisby, Red salad bowl and Mikola). Information from the NIAB(National Institute for Agricultural Botany) and the seed company (Enzazaden) as regards the different fungal susceptibilities of the different varieties shows that there is considerable variation. (Cresswell, 2004; Day, 2004). As it is, the lack of data means that we could not see whether there was a differential effect of the healer on the different varieties.

Farmer's Assessment of the bad growth of the crop

At the end of May the grower expressed great concern that the germination rate was very bad. She stated that the people working on the farm were really troubled about it and thought that it was something to do with the research. She was concerned that there were no plants growing outside of the experiment: a control of the controls so to speak. By the third week in June it was decided to only run 7 trials.

In the third week in July, the grower talked again about how "spooked" they all were, and that one of the people working on the farm said he "had never seen anything like it before." During the next week the grower talked with other farmers who offered various "psychic" or "energy" reasons for the problem. The following week the grower decided to terminate the experiment. Thus the assessment of the people working on the farm about the crop failure was that it was due to an "energy" or psychic effect, caused by the team of people coming on to the farm to do the research upsetting the delicate balance of energy on the farm and so affecting the germination of the lettuce seed.

Independent Assessments of possible causes of bad growth of plants

In parapsychology it is really important to look at all the possible reasons why something has occurred before accepting a psychic hypothesis. This is essential in the case of South farm where there was such a dramatic loss of plants whilst they were in the trays, of an order rarely seen in psychical research. Accordingly SRD consulted two organic crop research organisations, the Organic Advisory Service and the Henry Doubleday Research Association, keeping the identity of the farm confidential, and asked them if they would be able to help with information regarding problems with lettuce germination and growth. The raw data, concerning numbers of plants germinated, planted out and harvested, and photographs of the plants in trays and in the field, were sent to them. The Organic Advisory Service (Hitchings, 2003) gave a full report the main points of which are as follows:

The first point to note is that the independent consultants consider that the actual germination of the seeds was within the normal range that one expects for lettuces.

The second point is the possibility of a heater in the greenhouse affecting the plants. The weather at that time was unseasonably cold and wet and it is possible that a heater was being used at night, though it was not noticeably in use during the day. At no time did SRD notice that the greenhouse was hot, so this possibility seems a less likely factor.

The third point is that of inconsistency in watering. This had already been noticed by SRD and was something she had discussed with the person in charge of sowing the seeds. It was explained that the hose did not have a sprinkler, so that it was very difficult to give equal water to the plants and some got waterlogged whilst others were left dry. The photographs show this unequal watering very clearly.

The fourth point is that most of the failure occurred whilst in the greenhouse after germination and before planting out. Though there was also loss after planting out, it is considered to be in the acceptable range. The report mentions that seedlings that have survived the stress of over- or under-watering may not develop well afterwards, and this point needs to be taken into consideration also.

Unfortunately, as remarked by the grower, the whole of the lettuce crop was used in the research, and this meant that there was no outside control as there had been at Radford Mill, where the research lettuces were only a small part of the overall crop. Thus, whilst the farmer's assessment that the problem was due to a new group of people upsetting the "energy" on the farm, there are several other possibilities that need to be considered before this psychic reason for the loss of lettuces can be considered the valid option. It is possible that the farmer's assessment is an example of "fear of psi" manifesting.

Conclusion

For three years field trials have been run on organic farms in England in which a healer has directly enhanced the seed of the lettuce crop. Every year significant results have occurred. This paper shows that, even under adverse conditions, the effect of the healer is noticeable. As organic farming is becoming more and more viable and popular, the possibility of a healer working with seed companies to boost production is certainly to be considered.

REFERENCES

Benor, D.J. (1993) *Healing Research: Vol. 1*, Helix Press.

Benor, D.J. (2004) *Healing Research: Vol. II., Consciousness, Bioenergy and Healing: Self-Healing and Energy Medicinefor the 21st Century*, Wholistic Healing Publications.

Cresswell, A. (2004). Personal communication.

Day, M. (2004). Personal communication.

Hitchings, R. (2003). Personal communication.

Roney-Dougal, S.M. & Solfvin, J. (2002a). Field Study of enhancement effect on lettuce seeds - their germination rate, growth and health. *Journal of the Society of Psychical Research*, *66*, 129-143.

Roney-Dougal, S.M. & Solfvin, J.(2002b). Field study of an enhancement effect on lettuce seeds: A replication study, *Paper presented at the 45th Annual Convention of the Paraposychological Association, Paris, France, August, 2002.*

Roney-Dougal, S.M. & Solfvin, J. (2003) Field Study of an Enhancement Effect on Lettuce Seeds: A replication study. *Journal of Parapsychology, in press.*

Schwartz, S.A. & DeMattei, R. (1990) Infrared spectra alteration in water proximate to the palms of therapeutic practitioners. *Subtle Energies, 1*, 43-72.

Address for correspondence: Serena M. Roney-Dougal,
Psi Research Centre, 14 Selwood Rd.,
Glastonbury BA6 8HN, UK.
E-mail: serenard1@yahoo.co.uk

THE PRECOGNITIVE HABITUATION EFFECT: AN ADAPTATION USING SPIDER STIMULI

Louie Savva, Robert Child & Matthew D. Smith
Parapsychology Research Group, Department of Psychology
Liverpool Hope University College

ABSTRACT

There has been a recent trend in precognition research to examine established conventional psychological paradigms for temporally reversed effects. The precognitive habituation (PH) effect is a newly emerging paradigm based upon a temporally reversed mere exposure (ME) study. Where in conventional psychology the ME effect involves exposing stimuli to participants and then measuring liking for it, the PH effect involves the opposite procedure. Participants are presented with a pair of photographs and are then asked to make a preference choice between the two. Previous research has argued that precognitive exposure to one target over another, results in diminished arousal and that negatively arousing targets are made less negative. The PH hypothesis is that negatively arousing targets will be preferred over the non-targets and that no effect (or a precognitive boredom effect) is expected on non-arousing (low-affect) trial. In this study, we sought to conceptually replicate and extend Bem's (2003) findings by using less ethically problematic images. In the Bem studies, many of the images were disturbing (graphic images of gun shot victims etc) and Bem reported that some participants did make attempts to avoid looking at the stimuli through closing of the eyes or averting the gaze. To help circumvent the problem of showing potentially disturbing stimuli to participants, the authors sought to replace the negatively arousing images used by Bem with pictures of spiders. This was based on the work of Savva and French (2001; 2002), where spider pictures had been used to replace more disturbing images in a number of paradigms, including the presentiment effect. Fifty participants contributed to the current study and provided a self-report measure of spider fear, where 25 were categorised as 'spider fearing' and 25 as 'no-spider fear'. The overall hit rates obtained by the different fear groups were not significant but the data did suggest that an analogous effect was to be found. A significant difference was found for the mean number of hits obtained on the spider stimuli versus the low-affect pictures and only for the spider fearing group. This effect may be interpreted as a precognitive habituation or a temporally reversed mere-exposure effect. The results are discussed and future direction of research suggested.

INTRODUCTION

Within contemporary precognition research there seems to be an increasing interest to investigate what can best be termed "time-reversed effects" (TREs). That is, to take a conventional psychological paradigm and examine it for potential psi influences from the future. For example, Radin (1997) highlighted Klintman's (1983; 1984) work as an example of one of these TREs. Klintman was interested in a variation of the Stroop (1935) paradigm, where cognitive interference affects the processing of coloured names and words. In the standard Stroop paradigm, participants take longer to name the ink colour of a colour word if the ink colour and word meaning are incongruent (e.g. "red" written in blue ink) compared to when they are congruent (e.g. "red" in red ink). However due to unexpected variation in the data, Klintman reanalysed his results and found a stroop-like effect that was reversed in the temporal dimension; hence Klintman termed it time-reversed interference (TRI). Although there have been a number of studies investigating the TRI effect, the overall conclusions from the paradigm are mixed (e.g. Camfferman, 1987; Radin & May, 2000; Savva & French, 2002).

The presentiment effect is another example of a TRE, in that it is essentially temporally reversed arousal to emotional stimuli, not seen with low affect stimuli (Bierman & Radin 1998; 2000; Radin, 1996). In the early presentiment work, skin conductance provided a measure of arousal and anomalous anticipatory effects were claimed beyond normal expectancy effects. Bierman and Scholte (2002) chose to extend the presentiment effect by making use of an fMRI (Functional Magnetic Resonance Imaging) brain scan to highlight the brain's activities and monitor the cells that are responsible for perceptions and emotions. The experiment involved the participant being scanned while they were shown images in a random order. The images were either emotionally stimulating (these were violent or negative pictures) or neutral. Results were similar to the earlier presentiment effect studies and the presentiment effect continues to be an important area of precognition research.

Furthermore, Bem (2003) has recently reported positive findings investigating another TRE. This involved a variation upon a technique called the mere exposure (ME) effect. Conventional ME theories state that the more humans (or animals) are exposed to a particular stimulus the more they will like it. The large amount of research on mere exposure has established this as a well recognised phenomenon (Bornstein, 1989).

Bem (2003) hypothesised that the stimuli that would be most effective in a ME study would be negative and highly arousing, and highlighted the lack of ME research using such stimuli. It is also clear that ME takes effect mostly when the stimuli are shown subliminally, suggesting that the effect is taking place at an unconscious level. Using paradigms like the presentiment as an influence, Bem (2003) decided to test for a temporally reversed ME effect, which he referred to as precognitive habituation (PH). He devised a protocol to test for precognition, by conducting a ME study backwards. A trial in a PH experiment involves the participant being shown a pair of pictures and being asked to choose which picture they prefer. Following this the computer uses a pseudo-random procedure to select one of the two displayed pictures as the target and is then presented to the participant. The PH hypothesis is that the repeated flashing of the target image will weaken the arousal that it would have produced. Bem argues that will lead the negative pictures to be less negative and the positive pictures to be less positive. It is important to note that the two displayed images are matched for their arousal type (positive or negative). He found that the highly negatively arousing pictures produced a significant psi effect with participants preferring the randomly selected target picture more than would be expected by chance. A similar, although reversed, effect was found with highly positively arousing pictures. No such significance was found for low-affect stimuli.

Since Bem's precognitive habituation effect had reportedly produced a significant hit rate comparable to other popular parapsychological paradigms, the authors felt that it was worthwhile attempting a replication. Bem had taken stimuli from the International Affective Picture Set (IAPS, Lang and Greenwald, 1993) and supplemented these with images taken from the Internet. These pictures fell into three categories; the first was low-affect (e.g., coastlines), the second were negative images (e.g., pictures of corpses) and the third were erotic images. On viewing the highly negative and highly positive images used by Bem, however, it was deemed that these may prove to be ethically problematic in terms of attempting a replcation of the PH effect. Indeed Bem reported that in one PH study, female participants were recorded shutting their eyes, or averting their glance, thus reducing the PH effect. It was therefore concluded that alternative stimuli be sought.

One of the authors (LS) had already conducted presentiment research where conventional stimuli (such as that found in the Bem study) had been replaced with spider related stimuli. Savva and French (2001) argued that pictures of spiders would potentially induce in those afraid of spiders a similar reaction as the violent images would in 'normal' participants. By replacing the negative pictures with spider stimuli, and by omitting the erotic pictures, any ethical concerns we had with conducting a PH study would be resolved.

Savva and French have also replaced conventional stimuli with spider stimuli in a number of other paradigms, including the time-reversed interference effect, with varying degrees of success (Savva & French, 2002). The fear of spiders is one of the most common phobias in Western society (Cornelius & Averill, 1983; Kirkpatrick, 1984). It has been argued that the phobia stems from evolutionary selection due to some spiders being venomous (Seligman, 1971; Öhman, 1986). Davey (1994) highlights the fact that many studies (e.g. McNally, 1987) have carried out a research that linked pictures of spiders with a mild electric shock and have found that the fear which becomes conditioned to spiders is significantly more resistant to extinction than the fear of low-affect (or 'fear-irrelevant') images. In his own paper, Bem also suggests that a precognitive ability to detect the emotionally arousing images may be related to an evolutionary adaptation for psi.

Thus, based on the findings reported by Bem (2003), it was hypothesised that those participants who had a fear of spiders would be more likely to select the target picture when the pictures were of spiders than when the pictures were low-affect images. No such difference was expected for non-spider fearing participants.

METHODS

Design

The study was a 2x2 between-subjects design with two independent variables: emotionality of stimuli (spider versus calm) and fear group (spider-fearing versus no fear group). [1]

Participants

There were 50 participants (23 male, 27 female) representing an opportunity sample recruited mainly from the undergraduate population at Liverpool Hope University College. Equal numbers of spider-fearing and no fear participants were recruited. Age ranged from 18 to36 (with a mean of 22.4 years and a standard deviation of 3.08). Participants were not made aware of the exact aims of the study, but were given enough information to make an informed choice on whether to participate in the study or not.

Materials

The participants were categorised with a fear of spiders questionnaire devised by Szymanski and O'Donohue (1993). This test is comprised of six yes-no type questions; with an answer of yes to any categorising the participant as spider-fearing, and those answering no to all as non-spider fearing.

The experimental program was run on a laptop computer (working under Windows XP at screen resolution 1024x768) using a program compiled by Daryl Bem. The images used by Bem (2003) were removed and replaced. All previous negative images were replaced with images of spiders, while many low-affect pictures were also added to replace some of the more arousing neutral pictures. Although we did not match pictures on any specific rating scales, pictures were wherever possible, matched on closeness of appearance. The full collection of pictures used in the study can be obtained from the researchers. A 'normal' (non-touchpad) mouse was also attached to the laptop, so the participants could choose how to respond to the stimuli.

Procedure

Participants were recruited individually and took part in the experiment in a location where they felt comfortable, often the participant's home, although other locations were sometimes used. All of these areas were dimly lit and had no other stimuli that could distract the participant during the experiment. No other people were present in the room with the participant when they were viewing the images. All participants were asked to switch off their mobile phones to stop any disturbances during the experiment.

Before beginning the experiment all participants were given instructions regarding what to expect during the task and asked to fill out a consent form and a copy of the participant response form (which contained the spider questionnaire). Upon consulting the answers to the questionnaire participants were either allowed to continue or advised that the group into which they were categorised (spider-fear or no fear)

[1] Although one anonymous referee pointed out that we are specifically interested in just the difference of the spider-fear group on the spider stimuli versus the low-affect stimuli and that the no fear group are ex hypothesi, the authors agreed that all collected data should be presented. This comment may however influence any future designs.

already contained enough participants. On the occasions in which the latter occurred, the participant was given the option to continue with the study in the knowledge that their results would not be used.

The precognitive habituation software first presents a screening test, which comprises of a 20-item questionnaire. Although the data was collected during the current research, it was not used in any subsequent analysis. The software then provides a "cool-down" period, where participants are provided with 5 minutes of relaxing sounds and images.

On each trial, participants were shown two pictures side by side and asked to indicate which one they like better by clicking on it with the mouse. They were then asked to watch the screen, as one of the pictures was flashed rapidly on the screen (12 subliminal exposures in all). There were 12 spider trials and a further 36 low-affect trials. At the end of these trials the experimenter debriefed the participants as to the aims of the study. The entire procedure lasted approximately 20-25 minutes.

The PH program supplied by Daryl Bem included a database program that collated the data and exported the required data to SPSS for further analysis.

RESULTS

Plotting the results in a bar chart (see figure 1) reveals that those who reported a fear of spiders did score better than those who reported no-fear. There also seems to be a difference between spider stimuli and the low-affect stimuli for the spider-fearful participants, not seen in the no-fear group.

The hit rate for the overall study is 51.3% for the spider pictures and is not significantly above chance ($t(49) = .739$, p = ns) and 48.1% for the low-affect stimuli ($t(49) = -1.58$, p = ns). This splits across the fear groups where the spider fear group obtained a hit rate of 54% on the spider stimuli ($t(24) = 1.70$, $p = .051$, one tailed) and a 48% hit rate on the low-affect pictures ($t(24) = -1.07$, p = ns). The no fear group obtained a hit rate of 49% on the spider pictures ($t(24) = -.35$, p = ns) and 48% on the low-affect pictures ($t(24) = -1.14$, p = ns). As such the hit rate is in the predicted direction but is not significantly different from mean chance expectation.

A 2x2 repeated measures ANOVA was conducted on the data. No within subjects main effect of type of stimuli (spider versus low-affect) was found ($F(1,48) = 2.23$, p = ns). No between subjects main effect of fear group was found ($F(1,48) = .017$, p = ns). A slightly suggestive interaction was found ($F(1,48) = 2.97$, $p = .091$. Two t-tests were carried out to explore this suggestive interaction further (Bonferroni corrected to .025 to reduce spurious claims of significance). A non-significant difference was found for the spider stimuli and low-affect stimuli hit rates for the no-fear group ($t(24) = -1.51$ p = ns). However a significant difference was found for the spider stimuli and low-affect stimuli hit rates for the spider-fear group ($t(24) = 2.48$, $p = 0.021$). This suggests that for the spider-fear group performance on the PH task seems to be significantly different when the stimuli are spider related than when the stimuli are low-affect.

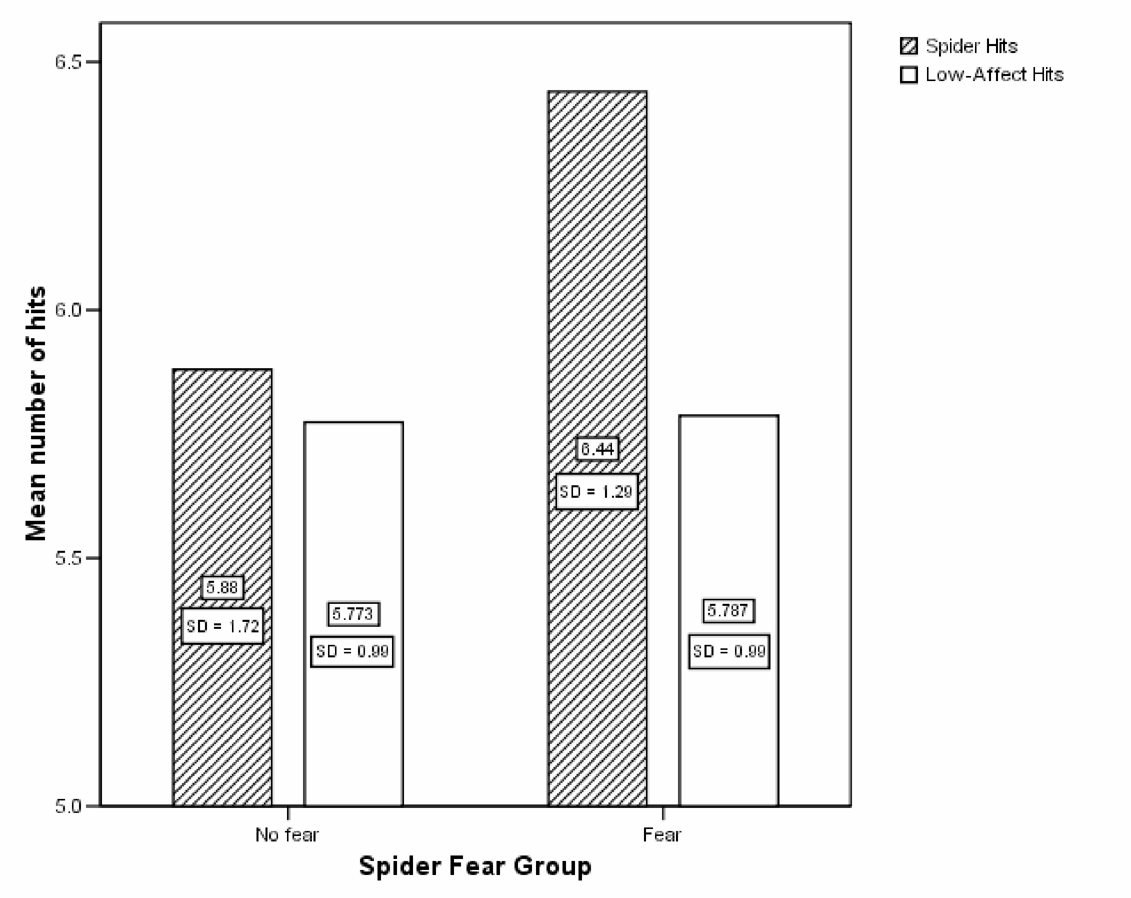


Fig. 1 Mean hits for both spider pictures and low-affect stimuli across the spider fear groups.

DISCUSSION

The results of this study are very encouraging. After replacing the stimuli found in earlier PH studies with spider-stimuli, it was found that participants who rated themselves as being afraid of spiders seemed to precognitively habituate to spider-stimuli and seemed to exhibit a precognitive boredom effect for low-affect stimuli. No psi effect was found in the no-fear group.

These results seem to suggest that participants who are afraid of spiders show preference to one picture over another matched picture, if they are then shown that picture in the future. This appears to be a temporally reversed ME effect. These participants also show a precognitive boredom effect for the low-affect stimuli, showing a tendency to select the picture that they will not be shown in the future. Although neither effect is very striking, the difference between the two effects is significant and not found in the no-fear group.

This adapted PH methodology therefore provides parapsychologists with an ethically less problematic tool than that developed by Bem. Whilst it is unlikely that the ethics committee at Liverpool Hope University College would have passed the PH experiment in its original form, it was more than happy to accept the version reported here. As such it provides a means of making use of what claims to be a replicable and simple methodology without subjecting participants to extremely unpleasant stimuli. Although some may argue that spider stimuli can be just as disturbing as Bem's original stimuli it was felt that, since participants would be informed that they would be presented with spider pictures, such stimuli, however unpleasant, were not as surprising as some of the images in Bem's stimuli set.

Future studies

Perhaps the PH effect's greatest strength is the ease with which it can be adapted and manipulated. Currently a straight replication is planned to follow up this initial study. If this is successful there are a number of adaptations that would be worthy of further investigation. Perhaps of interest to conventional psychologists and those believing the precognitive effect to be some kind of TRE is to conduct a normal mere-exposure effect study, combined with a PH study. If a similar level of mere-exposure is found for those rating themselves as spider-fearful to that found in the PH part of the study, it would certainly provide strength to the argument that it is a TRE and that it makes little difference what the temporal direction of the experiment is.

It would also be interesting to compare other fear groups, such as those afraid of snakes, and even to investigate clinical populations (i.e., patients who are diagnosed as extremely phobic of spiders). The spider-fear measure used in the study above is not very selective and even people who merely dislike spiders are rated as being afraid of spiders. By making the difference between spider-phobics and non-spider phobics more extreme it could be argued that a greater difference would be predicted.

Finally, the stimuli presentation method could be adapted to include some more life-like stimuli. Virtual reality is regularly used in the clinical treatment of phobias to provide extremely life-like (and therefore emotionally more realistic) exposure to negative stimuli. A virtual-reality based PH effect methodology could be developed which would produce a more naturalistic testing environment, whist maintaining the best elements of laboratory testing.

ACKNOWLEDGEMENTS

This paper is based on a study conducted by Robert Child for his undergraduate dissertation, under the supervision of Matthew Smith and Louie Savva. We would like to thank Professor Daryl Bem of Cornell University for his kind help in developing this experiment.

REFERENCES

Bem, D. J. (2003). Precognitive habituation: Replicable evidence for a process of anomalous cognition. *Proceedings of Presented Papers: The Parapsychological Association 46th Annual Convention*. 6-20.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin*, 106, 265-289.

Bierman, D. J., & Radin, D.I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills*, 84, 689-690.

Bierman, D. J., & Scholte, H. S. (2002). Anomalous anticipatory brain activation preceding exposure of emotional and neutral pictures. *The Parapsychological Association 45th Annual Convention, Proceedings of the Presented Papers*, 25-36.

Camfferman, G. (1987). Time reversed interference: A replication study. *European Journal of Parapsychology*, 7, 13-31.

Cornelius, R. R. & Averill J. R. (1983). Sex differences in fear of spiders. *Journal of Personality & Social Psychology*, 45, 377-383.

Davey, G. C. L. (1994). The disgusting spider: The role of disease and illness in the perpetuation of fear of spiders. *Society and Animals*, 3, 17-24.

Dijksterhuis, A., & Smith, P. K. (2002). Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. *Emotion*, 2, 203-214

Klintman, H. (1983). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? I. European Journal of Parapsychology, 5, 19-49.

Klintman, H. (1984). Is there a paranormal (precognitive) influence in certain types of perceptual sequences? II. European Journal of Parapsychology, 5, 125-140.

Kirkpatrick, D. R. (1984). Age, gender and patterns of common intense fears among adults. *Behaviour Research & Therapy*, 22, 141-150.

Lang, P. J., & Greenwald, M. K. (1993). International Affective Picture System Standardization Procedure and Results for Affective Judgments: Technical Reports 1A-1C: University of Florida Center for Research in Psychophysiology, Gainesville, FL.

McNally, R. J. (1987). Preparedness and phobias: A review. *Psychological Bulletin*, 101, 283-303.

Öhman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. Psychophysiology, 23, 123-145.

Radin, D. (1997). The Conscious Universe: The Scientific Truth of Psychic Phenomena. San Francisco, Harper.

Radin, D. & May, E. (2000). Evidence of a retrocausal effect in the human nervous system [World Wide Web]. Retrieved from http://www.boundaryinstitute.org/articles/tri2.pdf.

Savva, L., & French, C. C. (2001). *Investigating the presentiment effect as an adapted behaviour. Evolution and psi.* 25th International Conference of the Society for Psychical Research. Clare College, Cambridge, 14-16 September 2001.

Savva, L., & French, C. C. (2002). Is there time-reversed interference in Stroop- based tasks. *The Parapsychological Association 45th Annual Convention, Proceedings of the Presented Papers*, 194-205.

Seligman, M.E.P. (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307-320.

Szymanski, J., O'Donohue, W. (1995) Fear of spiders questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, 26, 1, 31-34

Address for correspondence: Matthew D. Smith,
Parapsychology Research Group, Department of
Psychology, Liverpool Hope University College,
Hope Park, Liverpool, L16 9JD, UK.
E-mail: smithm3@hope.ac.uk

INTERPERSONAL PSI: EXPLORING THE ROLE OF THE EXPERIMENTER AND THE EXPERIMENTAL CLIMATE IN A GANZFELD TELEPATHY TASK

Simon J. Sherwood[1], Chris A. Roe[1], Nicola J. Holt[1], & Stuart Wilson[2]

[1]*Centre for the Study of Anomalous Psychological Processes*

University College Northampton

[2] *Department of Psychology*

Queen Margaret University College, Edinburgh

ABSTRACT

The aim of this study was to examine the relationship between actual participants' pre-session perceptions of their mood and expectations of success, their rapport and the quality of the interactions with the experimenter, and the subsequent outcome of the ganzfeld telepathy sessions. A total of thirty-eight trials were conducted by three different experimenters (NH, CR, SS) using an automated free-response testing system. Participants consisted mainly of a sample of volunteers from the local community, staff and students at University College Northampton (UCN) and friends and acquaintances of the experimenters. Overall the mean z score based upon ratings of the target relative to the dummy video clips was slightly negative (mean $z = -0.015$, $SD = 0.81$) and did not differ significantly from zero (one-sample $t(37) = -0.11$, $p = 0.90$, two-tailed[1] $r = -0.002$). There were no significant differences in the mean z score across the three experimenters. Based upon the participants' and experimenters' ratings, it is clear that all those involved tended to be fairly relaxed and in a positive mood at the end of the pre-session briefing and perceived the quality of the interactions to have been warm, fairly spontaneous and positive. Neither the participants nor the experimenters rated their confidence that the session would be a success as being extremely high, though the experimenters' ratings tended to be higher than the participants'. The findings concerning potential links between participants' and experimenters' ratings of their pre-session personal feelings and expectations and their perceptions of the rapport and quality of interaction between all parties are somewhat unclear and in some cases, even contradictory, and there was no correction made for multiple analyses. Of all of the roles involved in this experiment, if there is any possible trend it is for the senders' ratings suggesting that trials were slightly more successful the more relaxed, optimistic, and confident of success they were and the greater the perceived rapport and quality of interaction with the experimenter and the receiver. There is also a suggestion ($r = -.332$, $p = .113$, two-tailed), albeit a non-significant one, that the more relaxed the experimenters feel before the session the higher the ratings given to the target clip at the end of the sessions. This seems to provide further support for the argument that more attention should be paid to pre-session preparation by the experimenters (e.g., Delanoy, 1997; Schlitz, 1987; Watt et al., 2002). Our quantitative approach has been somewhat unsuccessful and perhaps it is, by definition, unable to adequately tackle the complexity and richness of human verbal and non-verbal interactions. We support the endeavours of other researchers who have taken or recommended such a more qualitative approach (e.g., Schlitz, 1987; Smith, 2003a; Watt et al., 2002; White, 1997). More qualitative analyses of, for example, filmed interactions between clearly psi-inhibitory and psi-conducive experimenters (cf. Wiseman & Schlitz, 1997, 1999) may provide more reliable indicators of performance.

INTRODUCTION

Replication of positive research findings is a crucial issue within parapsychology and the apparent lack thereof is a common criticism. There seems to be a common belief that some experimenters (such as Charles Honorton, William Braud, Marilyn Schlitz and Kathy Dalton) are 'psi-conducive,' whereas others

[1] This prediction was originally one-tailed, but in the opposite direction to that found. The two-tailed probability is given for information only.

(such as James Crumbaugh, Donald West and John Beloff) appear to be 'psi-inhibitory' (cf. Irwin, 1999, pp. 84-86). Such consistent differences in findings could be partly due to the experimenter him/herself and/or partly due to the experimental environment and climate in which the research is conducted. White (1977, p. 273) noted that "The experimenter has been a neglected variable in parapsychological research ... [yet] ... there could hardly be a more significant area of investigation than the role of the experimenter". Gardner Murphy (1949) suggested that there is no such thing as a gifted participant as such, but rather how well a participant scores on a psi task depends on the person who does the testing and the nature of the experimental conditions. Smith (2003a) has summarised four possible explanations for such an experimenter effect: (1) experimenter error—where successful experimenters make motivated but unconscious errors during their experiments; (2) experimenter fraud; (3) experimenter-participants interaction; (4) experimenter psi.

A warm social ambience is considered to be very important in a ganzfeld context (Bem & Honorton, 1994; Dalton, 1997; see also Milton & Wiseman, 1999). The experimental environment and climate and the interactions between the experimenter and the participants are considered to be important factors (e.g., Delanoy, 1997; Schmeidler, 1988; White, 1976, 1977) but have not received much experimental attention in terms of establishing how important these actually are and what specific aspects are the greatest contributors and how these might be harnessed.

It is well-known in mainstream psychology that participants look to the experimenter for cues to what is expected of them – such cues are referred to as 'demand characteristics.' These cues might be picked up from the environment, what the experimenter says about the experiment and how this is expressed, and/or from the nonverbal body language and behaviour of the experimenter. Rosenthal (1966) has demonstrated how variations in experimenter expectancy or orientation can lead to quite different findings. Harris and Rosenthal (1985) later reported a meta-analysis of 135 subsequent experiments that broadly confirmed the impression that the expectancy of experimenters could produce an unexpected but significant effect upon study outcome. Successful experimenters in the field of parapsychology might be more highly motivated to obtain positive results (Rhine, 1948) and more likely to believe in the possibility of psi and therefore have greater expectations of success (Dalton, 1997; Schlitz, 1987; Schmeidler, 1988; Smith, 2003a; Parker, 1975; Taddonio, 1976; White, 1976, 1977; Wiseman & Schlitz, 1997, 1999). In a recent test of this hypothesis, based upon perceived ratings, Smith (2003b) found medium to large positive correlations between psi-conduciveness ratings and experimenters' belief in own psi ability, belief that psi is possible and belief that it can be demonstrated in a laboratory setting. Sharp and Clark (1937), using several experimenters who differed in their attitude towards psi, obtained results that seemed to relate experimenters' attitudes to participants' subsequent hit rates; Sharp (mean hit rate = 5.36, where MCE = 5.00) and Davidson (5.88) were positive towards the existence of psi, whereas Berger (4.86) was uncommitted, and Myers (4.30) was sceptical. However, typically for a quasi-experiment, it is difficult to know whether these experimenters may have differed from one another in other ways that might have influenced participant scoring.

The pre-experimental briefing session might be a key point during which the experimenter's attitudes and expectations are communicated and a rapport is or is not established with the participants (Smith, 2003a). Experimenter orientation and preparation also seems to be an important factor (Delanoy, 1997; Schlitz, 1987; Watt, Wiseman, & Schlitz, 2002) albeit one that has received little empirical investigation. Some experimenters might spend more time briefing and getting to know their participants and make more of an effort to establish rapport than others (e.g., see Smith, 2003a; Watt et al., 2002). Potentially psi-conducive practice might include making participants feel comfortable and valued with positive expectations and a motivation to succeed in the experimental session (Delanoy, 1997). Woodruff and Dale (1950) were the first to include measures of the experimenter-participant attitudes towards one another. Surprisingly, higher scores were associated with participants who gave lower ratings for experimenters. It should be noted however, that 'low' ratings were actually neutral rather than very negative, and may indicate that disinterest or antipathy is desirable. For Dale, there was a difference in participant scores such that those given higher liking ratings by her subsequently performed significantly better at the psi task. Parker, Millar and Beloff (1977) attempted to assess the relationship between the rapport established between the experimenter and participants (during the initial interactions involving the explanation of the study and experimental instructions) and subsequent ganzfeld performance. Rapport was assessed by blind independent judges who listened to tape-recordings of the interactions for the three experimenters. Although there were significant differences in the amount of rapport established by the different experimenters, rapport did not have a significant effect on performance.

Two recent EDA DMILS studies (remote-staring PK) found that a psi-proponent experimenter obtained significant findings whereas a psi-sceptical experimenter following the same procedure did not (Wiseman & Schlitz, 1997, 1999). The authors of these studies noted that two possible explanations for these differential results were that they had used their own psi abilities or had acted differently towards their participants prior to the start of the experiment. It is difficult to manipulate deliberately the manner in which experimenters interact with their participants but nevertheless there is some suggestive evidence that experimenters who attempt to be informal, friendly, and supportive might get more favourable results, or at least are perceived more favourably by their participants (Honorton, Ramsey, & Cabbibo, 1975; Pratt & Price, 1938; Schneider, Binder, & Walach, 2000; Watt & Baker, 2002).

Smith (2003a) summarised the results of studies (Edge & Farkash, 1982; Schmeidler & Maher, 1981) that have investigated the perceived personality and behavioural characteristics of 'psi-conducive' and 'psi-inhibitory' experimenters as follows:

> Supposedly 'psi- conducive' experimenters tended to be rated as more flexible, friendly, free, likeable, warm, enthusiastic and playful whilst 'psi-inhibitory' experimenters were rated as more rigid, cold, overconfident, tense, irritable, egoistic and unfriendly. An attempted replication found 'psi-conducive' experimenters to be perceived as more active, nervous and enthusiastic with 'psi-inhibitory' experimenters being perceived as more poised, egoistic, cold and confident (Edge & Farkash, 1992, pp. 171-2). Thus, although in both studies there was a good deal of variance between judge's ratings, there would appear to be some agreement among potential participants that 'successful' psi experimenters come across as more enthusiastic, warmer and less egoistic than do their less successful counterparts. (p. 77).

However, an earlier study by Parker (1975), in which experimenters, later classified as either 'psi-conducive' or 'psi-inhibitory' by three judges, completed a 16 Personality Factor (16PF) measure, found no significant differences between these two groups in terms of extraversion, warmth and sociability, confidence adequacy or tenderness and sensitivity. This study differed from the later aforementioned studies in that the experimenters rated their own personality characteristics. Perhaps it is participants' perceptions of the experimenter rather than the experimenter's self-perception which is important.

We plan to look at the relationship between actual participants pre-session perceptions of their mood and expectations of success, their rapport and the quality of the interactions with the experimenter, and the subsequent outcome of the sessions. In terms of the participants' ratings, it is vital that participants feel that they can be completely honest about their perceptions of their experimenter and thus it is important that the experimenter remains blind to the ratings of individual participants. In our study, participants placed their completed unseen ratings in a sealed envelope that was posted to an independent researcher (Wilson) who would analyse these data. At no time would the experimenters see the interaction ratings from the individual participants. We agree with Smith (2003a) that a qualitative analysis of videotaped interactions between the experimenters and participants during this briefing session might shed further light on which aspects might be important and thus we planned to videotape this as part of this study, although these data are yet to be analysed and will not be reported here. Due to the difficulties associated with trying to manipulate experimenters' behaviour deliberately and the artificial nature of this, we decided simply to videotape our experimenters' natural behaviour.

METHOD

Design

This study adopted a correlational design in which experimenter, sender and receiver ratings of their current pre-session mood, relaxation-tension, pessimism-optimism, confidence of success in the experiment, and five measures of the rapport/interaction between the experimenter and the participants, were correlated with the measure of performance at the telepathy ESP task. The measure of performance was pre-specified as the *z* score of the target clip's similarity rating. A series of three pilot sessions and 40 trials were planned. However, a computer hardware failure after trial 38 resulted in the premature completion of data collection for this experiment. Two sessions were run but were excluded from the study: one because the sender failed to turn up for the session so it was run as a clairvoyance trial; one because the receiver needed to visit the bathroom during the mentation period and thus violated security protocols.

Apparatus and Materials

This study used an automated ganzfeld computer system developed by Dr Paul Stevens and written in Microsoft Visual Basic v5 that presented video material via the API for Media Player v7. Video clips are stored digitally as MPEG files, labelled 1a, 1b, 1c etc. Three separate monitors for the experimenter, sender and receiver are controlled by the experimenter's PC. Security measures within the program lock the experimenter out of the system completely during a session so that it is not possible to switch to another application or access the computer except by aborting the session.

The target set consists of 116 minute-long digital video clips arranged in 29 sets of 4. These were the same clips used in our previous ganzfeld study (Roe, Sherwood, & Holt, 2003) but some of the sets had been re-arranged to make them more orthogonal. These digital clips have mainly been produced at UCN, drawn from popular television programmes and commercial films, although some have been taken from the pool previously used at Edinburgh. Copies of the target pool are available on CD from the first author on request. Randomisation is achieved using the Visual Basic pseudo-random algorithm (rnd), having seeded it using the timer at the start of the program (RANDOMIZE TIMER). Once the "Start" button has been pressed, the computer first selects a target set, then selects one of the 4 clips within that set. The order of presentation of the four clips at judging is similarly randomised.

All trials were completed using specialist facilities in the Psychology Building at UCN. The receiver room is sound attenuated and is separated from a public corridor by two lockable doors. The sender's and receiver's rooms are separated by approximately 38m. A security camera is located outside the sender's room so that any activity there can be monitored by the experimenter and automatically video recorded.

The Participant Information Form (PIF) is a 56-item measure that was constructed for general use with parapsychological research at University College Northampton and includes questions concerning biographical and contact details (11-items); religious and parapsychological background (5 items); computer experience (2 items); practice of mental/physical disciplines (2 items); belief in luck (2 items); clumsiness and punctuality (2 items); competitiveness (1 item); absorption (2 items); sleep and dreams (4 items); imagination and fantasy-proneness (3 items); creativity (2 items); and physical and mental health (1 item). The remaining items relate specifically to knowledge, belief and experience of anomalous phenomena including telepathy, clairvoyance, precognition, psychokinesis, 'communication with the dead' and out of body experiences (19 items). The form concludes with questions about hypnagogic/hypnopompic experiences in a range of modalities (10 items) and an open question inviting descriptions of personal anomalous sleep-related experiences. Copies of all in-house measures are available from the first author on request.

Participants also completed the short extraversion and neuroticism subscales of the EPQ-R (Eysenck, Eysenck, & Barrett, 1985). Each subscale has 24 items with a dichotomous yes/no response format. The 18-item Australian Sheep-Goat Scale (ASGS, Thalbourne & Delin, 1993), with a 5-point Likert scale ranging from strongly agree to strongly disagree, was also completed.

At the end of the pre-session briefing, the experimenters and participants completed a short Interaction Questionnaire containing nine questions, concerning their personal feelings and expectations and perceptions of the quality of the interactions between experimenter and participants, which they were required to answer by giving ratings on 7-point scales:

How would you rate your current mood? (Negative-Positive)
How do you feel at this moment? (Relaxed-Tense)

How do you feel about the prospect of participating in this experiment? (Pessimistic-Optimistic)
How confident are you that today's experiment will be a success? (Not at all confident-Extremely confident)
How would you describe the quality of rapport that you have with the other Participant? (Extremely poor-Extremely good)
How would you describe the quality of rapport that you have with the Experimenter? (Extremely poor-Extremely good)
How would you rate the quality of the interaction between experimenter and participants? (Very cold-Very warm, 'Rehearsed'-Spontaneous, Very negative-Very positive)

Post-ganzfeld measures included a Sender Strategy Questionnaire that asked about the type of sending strategies used, whether this was active or passive, holistic or atomistic, focused on target clip or on the receiver, realistic or associative, and continuous or episodic. A Receiver Questionnaire asked about the receiver's experience.

Participants

An opportunity sampling method was used to draw 38 pairs of participants (mean age of senders = 38.8 years [range = 19 – 71], 9 males and 29 females; mean age of receivers = 39.3 years [range = 19 – 72], 24 males and 15 females). These mainly consisted of friends and acquaintances of the experimenters, and staff and students at UCN, although attempts were made to recruit participants from the wider community using posters and media appeals. Participants were not selected on the basis of prior belief or experiences, or personality and attitudinal dimensions that may predict psi performance (although such variables were measured). In most instances, each participant provided his or her own sender but in some circumstances individuals were paired up by the experimenters. Lab personnel did not serve as participants. The mean ASGS belief score for receivers in this sample was 58.1 (std dev = 14.4) and for senders was 50.5 (15.8). The receivers' belief mean is slightly above the theoretical mean for the scale (mean = 54), suggesting that the sample are moderate believers; the senders' mean is slightly below the mean suggesting that they are moderately sceptical. Among this sample, eight senders and twelve receivers had previously participated in formal parapsychological studies; nine senders and 16 receivers had previously participated in casual testing; 18 senders and 22 receivers had practised a mental discipline such as meditation; and 14 senders and 20 receivers had practised a physical or spiritual regimen such as yoga or tai chi. All three authors acted as experimenters in the running of trials, with NH (E1) conducting 14 trials, CR (E2) 13 trials and SS (E3) 11 trials.

Procedure

Potential participants were sent an information sheet illustrated with photographs that described the nature of the study. This provided a rationale for the ganzfeld paradigm, outlined the stages of the experimental procedure, focusing on the roles of the experimenter, sender and receiver. Thus participants were made fully aware of all aspects of the experiment so that those who were not comfortable with the procedure had the opportunity to withdraw from the study. Prior to the trial, participants (senders and receivers) completed a battery of measures. A video player was set to record the input from the security video camera as the experimenter prepared for the session and continued recording until after the session was over. Participants were greeted on arrival and escorted to a reception room that had been specially prepared with comfortable chairs, a coffee table, rugs and curtains so as to make participants feel as comfortable and relaxed as possible prior to the trial. This room contained two opposite wall-mounted cameras, which focused upon the seated experimenter and participants, respectively and a wall-mounted microphone. Signals from the cameras and microphone were recorded on an independent videorecorder in the experimenter's room. Participants were reminded about the cameras and potential recording of the briefing session only and asked to consent to this before entering the room (all participants gave their consent). As recommended by Honorton et al. (1990), experimenters encouraged an informal, friendly and positive atmosphere, discussing the procedure and answering any questions arising while sharing refreshments. The briefing included the background and rationale for ganzfeld ESP research plus a description of the procedure (see Honorton et al., 1990). Participants were then given a guided tour of the facility, as recommended by Honorton et al. (1990) and Dalton (1997), as the roles of sender and receiver were again explained.

With the assistance of the sender, the experimenter prepared the receiver for the ganzfeld and wished them success. The receiver was seated in a reclining chair and encouraged to relax. They were invited to remove their shoes and cover themselves with a blanket if desired. The receiver wore headphones that had a microphone attached through which they could communicate with the experimenter and be heard by the sender. Halved ping-pong balls were placed over their eyes and held secure with micropore tape. A red light was shone on the receiver's face, positioned immediately in front of them at a distance that was comfortable for them (typically one metre). The receiver was then locked in the room (unless they were uncomfortable with this) and the sender was guided back to their room.

Once the experimenter had returned to the experimenter's room and established contact with the receiver the trial commenced. The receiver began by listening to and following a series of progressive relaxation instructions. The sender watched a randomly-selected video clip that was played fifteen times with one minute intervals between plays (to allow them a break and/or time in which they could focus upon sending details of the clip without being distracted by it). Drawing materials were provided for the sender should they wish to sketch elements of the target clip during these 'quiet' periods. During this thirty-minute sending/mentation period the receiver listened to white noise being played through their headphones and reported on any impressions or sensations that they experienced. The experimenter listened to the receiver's mentation via headphones from the experimenter's room and took notes. The sender could also hear any comments made by the receiver during the mentation period.

Following the mentation period, the experimenter read the receiver's mentation back to them and asked if there was anything further that they would like to add or elaborate upon. The receiver was then asked a series of questions regarding their experiences in the ganzfeld. Simultaneously, the sender completed a questionnaire concerning their interaction with the target and sending strategies employed. At the judging stage the receiver was asked to remove their eye-shields but was encouraged to remain in a relaxed state as they watched four video clips, giving each one a rating (0-99) in terms of its similarity to their ganzfeld mentation. After viewing all four clips, they were able to view any or all of them as many times as they wished and to alter their ratings if necessary. The sender was able to listen to the clip soundtracks and the interaction between the receiver and experimenter during the judging stage, but did not view the dummy clips. Once the receiver was satisfied with their ratings, these were confirmed and saved as a permanent record. Only after the data were saved was the target clip revealed and replayed. The experimenter then collected the sender and accompanied him/her to the receiver's room where they convened with the receiver for a discussion and debriefing session. A copy of the trial data record was printed off and signed by all parties to confirm the details of the session.

RESULTS

Our primary psi measure was pre-specified as the *z* score of the target clip's similarity rating. To facilitate future meta-analyses, Table 1 also gives the number of direct hits and ranks data.

Table 1: Target rank frequencies

	Rank					
	1	2	3	4	Rating mean *z*	Rating SD *z*
NH (E1)	4 (28.6%)	5 (35.7%)	3 (21.4%)	2 (14.3%)	0.048	0.74
CR (E2)	3 (23.1%)	5 (38.4%)	1 (7.7%)	4 (30.8%)	0.002	0.96
SS (E3)	1 (9.1%)	4 (36.4%)	2 (18.1%)	4 (36.4%)	- 0.115	0.77
Overall	8 (21.1%)	14 (36.8%)	6 (15.8%)	10 (26.3%)	- 0.015	0.81

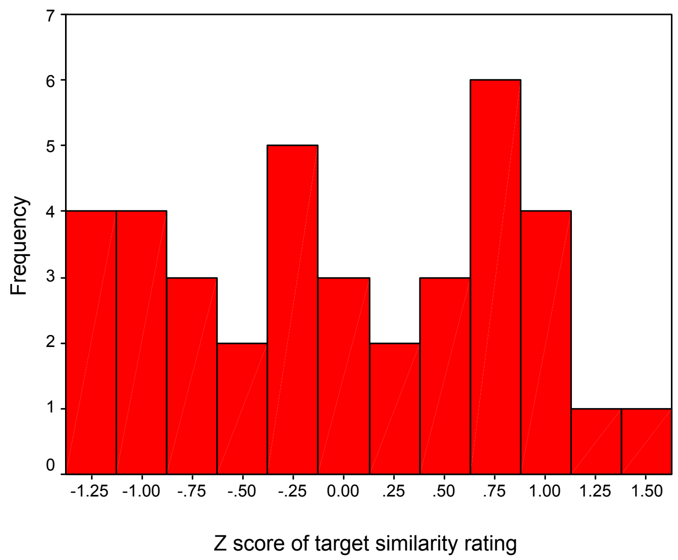


Figure 1: Frequency histogram of target similarity rating z scores

To assess our prediction that participants would award a similarity rating to the target that was higher than the average rating for the three dummy clips, *z* scores were calculated. The distribution of *z* scores is given in Figure 1. The hypothesis was not supported. The overall mean *z* score was actually slightly negative (mean $z = -0.015$, SD = 0.81) but did not differ significantly from zero (one-sample $t(37) = -0.11$, $p = 0.90$, two-tailed[1] $r = -0.002$). A Kruskal-Wallis test revealed no significant difference in the *z*-score for the target rating between trials conducted by the three experimenters ($\chi^2 = 0.20$, df = 2, p = 0.906).

Pre-session perceptions of senders, receivers and experimenters

Table 2: Mean receivers' ratings presented separately per experimenter

Receiver All ratings on scale 1-7	**E 1** Max N = 13	**E 2** Max N = 12	**E 3** Max N = 11	**Overall** Max N = 36
Current mood (Negative-Positive)	5.46	5.50	5.82	5.58
Relaxed-Tense	2.92	3.00	2.27	2.75
Pessimistic-Optimistic	5.08	5.33	5.60	5.31
Confidence of success (Not at all-Extremely)	4.58	4.33	4.82	4.57
Quality of rapport with other participant (Extremely poor-Extremely good)	5.69	6.08	6.00	5.92
Quality of rapport with experimenter (Extremely poor-Extremely good)	6.23	5.25	5.55	5.69
Quality of interaction between				

[1] This prediction was originally one-tailed, but in the opposite direction to that found. The two-tailed probability is given for information only.

experimenter and participants				
Very Cold-Very Warm	6.23	6.17	6.27	6.22
Rehearsed-Spontaneous	5.85	5.58	4.64	5.39
Very Negative-Very Positive	6.08	6.08	6.18	6.11

Table 3: Mean senders' ratings presented separately per experimenter

Sender All ratings on scale 1-7	**E 1** Max N = 13	**E 2** Max N = 12	**E 3** Max N = 11	**Overall** Max N = 36
Current mood (Negative-Positive)	5.46	5.25	6.18	5.61
Relaxed-Tense	2.54	2.83	2.27	2.56
Pessimistic-Optimistic	5.31	5.00	4.91	5.08
Confidence of success (Not at all-Extremely)	4.23	4.42	4.09	4.25
Quality of rapport with other participant (Extremely poor-Extremely good)	5.92	6.25	6.36	6.17
Quality of rapport with experimenter (Extremely poor-Extremely good)	5.69	5.25	5.82	5.58
Quality of interaction between experimenter and participants				
Very Cold-Very Warm	6.31	6.17	6.27	6.25
Rehearsed-Spontaneous	5.38	5.00	4.70	5.06
Very Negative-Very Positive	6.08	6.17	6.27	6.17

Table 4: Mean experimenters'ratings presented separately per experimenter

Experimenter All ratings on scale 1-7	**E 1** Max N = 13	**E 2**	**E 3** Max N = 11	**Overall** Max N = 24
Current mood (Negative-Positive)	4.08	~	5.82	4.88
Relaxed-Tense	3.54	~	2.82	3.21
Pessimistic-Optimistic	4.54	~	5.29	4.80
Confidence of success (Not at all-Extremely)	4.77	~	4.82	4.79
Quality of rapport with other participant (Extremely poor-Extremely good)	~	~	~	~
Quality of rapport with experimenter (Extremely poor-Extremely good)	~	~	~	~
Quality of interaction between experimenter and participants				
Very Cold-Very Warm	4.85	~	5.45	5.13
Rehearsed-Spontaneous	4.85	~	4.36	4.63
Very Negative-Very Positive	4.77	~	5.64	5.17

Note that due to an oversight, experimenter 2 (E2) did not complete any experimenter ratings prior to the session, only sender and receiver ratings. Ratings from one pair of participants could not be analysed because they had failed to indicate which of them was the sender and receiver. One set of ratings was not received by the independent data analyser.

Personal feelings and expectations

Across all three experimenters, both senders and receivers tended to be in a positive mood, fairly relaxed, optimistic about the prospect of participating in the experiment but not particularly confident that the session would be a success (see Tables 2-4). In fact, confidence of success was the aspect that was most lacking in the participants. The two experimenters, for whom data are available, were slightly more confident of success than their participants.

Perceived rapport and quality of interaction

Across all three experimenters, both senders and receivers tended to perceive the interaction between the experimenter and the participants as being warm, fairly spontaneous and positive (see Tables 2-4). The two experimenters, for whom data are available, had similar perceptions albeit not quite as positive.

The relationship between the pre-session perceptions of senders, receivers and experimenters and subsequent ESP task performance

Spearman's rho correlation coefficients were calculated between each of the receiver, sender, and experimenter ratings and the trial outcome, in terms of the *z*-score of the rating given to the target relative to the others in the set, for each of the three experimenters. The findings were mixed across the three experimenters, which suggests that the data should perhaps not be combined in order to calculate overall correlation coefficients, but these are reported for information. There was no correction for multiple analyses.

Table 5: Receivers' ratings per experimenter – Spearman's correlations with z-score target rating (two-tailed)

Receiver (p value)	**E 1** Max N = 13	**E 2** Max N = 12	**E 3** Max N = 11	**Overall** Max N = 36
Current mood (Negative-Positive)	.097 (.752)	- .568 (.054)	- .441 (.174)	- .335* (.046)
Relaxed-Tense	.281 (.352)	- .198 (.537)	.269 (.424)	.150 (.382)
Pessimistic-Optimistic	- .024 (.938)	.153 (.635)	- .281 (.431)	- .038 (.830)
Confidence of success (Not at all-Extremely)	.493 (.104)	- .118 (.715)	.183 (.589)	.184 (.289)
Quality of rapport with other participant (Extremely poor-Extremely good)	- .082 (.791)	- .547 (.066)	- .172 (.614)	- .289 (.088)
Quality of rapport with experimenter (Extremely poor-Extremely good)	.212 (.487)	- .467 (.126)	- .526 (.096)	- .285 (.092)
Quality of interaction between experimenter and participants				
Very Cold-Very Warm	.060 (.846)	- .366 (.241)	.074 (.830)	- .125 (.469)

Rehearsed-Spontaneous	.346 (.248)	- .144 (.654)	- .262 (.437)	- .063 (.713)
Very Negative-Very Positive	.425 (.147)	- .030 (.927)	- .083 (.807)	.076 (.660)

Table 6: Senders' ratings per experimenter – Spearman's correlations with z-score target rating (two-tailed)

Sender	**E 1** Max N = 13	**E 2** Max N = 12	**E 3** Max N = 11	**Overall** Max N = 36
Current mood (Negative-Positive)	.068 (.825)	- .190 (.555)	.353 (.287)	.033 (.847)
Relaxed-Tense	- .246 (.419)	- .416 (.179)	- .171 (.616)	- .268 (.114)
Pessimistic-Optimistic	.358 (.230)	.370 (.236)	.664* (.026)	.432** (.008)
Confidence of success (Not at all-Extremely)	.168 (.584)	.571 (.052)	.511 (.108)	.398* (.016)
Quality of rapport with other participant (Extremely poor-Extremely good)	- .052 (.865)	.247 (.439)	.178 (.600)	.147 (.391)
Quality of rapport with experimenter (Extremely poor-Extremely good)	.028 (.927)	.480 (.114)	.100 (.771)	.178 (.298)
Quality of interaction between experimenter and participants				
Very Cold-Very Warm	.167 (.586)	.190 (.554)	.427 (.191)	.227 (.183)
Rehearsed-Spontaneous	.377 (.204)	.161 (.617)	.459 (.182)	.309 (.071)
Very Negative-Very Positive	.331 (.270)	- .168 (.602)	.044 (.897)	.077 (.654)

Table 7: Experimenters' ratings per experimenter – Spearman's correlations with z-score target rating (two-tailed)

Experimenter	**E 1** Max N = 13	**E 2** Max N = 12	**E 3** Max N = 11	**Overall** Max N = 24
Current mood (Negative-Positive)	- .015 (.962)	~	.338 (.309)	.083 (.701)
Relaxed-Tense	- .517 (.070)	~	- .325 (.329)	- .332 (.113)
Pessimistic-Optimistic	.223 (.464)	~	.075 (.873)	.105 (.659)
Confidence of success (Not at all-Extremely)	.086 (.780)	~	.093 (.785)	.080 (.709)
Quality of rapport with other participant (Extremely poor-Extremely good)	.595* (.032)	~	- .017 (.959)	.256 (.226)
Quality of rapport with experimenter (Extremely poor-Extremely good)	~	~	~	~
Quality of interaction between experimenter and participants				
Very Cold-Very Warm	.406 (.169)	~	.207 (.542)	.133 (.535)
Rehearsed-Spontaneous	.038 (.903)	~	.149 (.662)	.091 (.674)
Very Negative-Very Positive	.096 (.756)	~	.268 (.425)	.096 (.655)

Personal feelings and expectations

There were no consistent findings across all three experimenters for the receivers' ratings (see Table 5). For experimenters 2 (E2) and 3 (E3), more negative receiver moods were associated with higher target ratings (r = - .568, - .441, respectively). For E1 and E3, higher ratings of receiver tension were associated with higher target ratings (r = .281, .269, respectively) but the reverse was true for E2 (r = - .198).

Findings were mixed in terms of receivers' optimism ratings and confidence of success, but it is interesting to note a large positive correlation (r = .493) between confidence ratings and target ratings *z*-scores for E1.

The senders' ratings (see Table 6) were the most consistent of the three roles involved with the experiment, at least in terms of the findings associated with their personal feelings and expectations. For all three experimenters, more relaxed, more optimistic and more confident senders tended to be associated with higher target ratings, particularly for E2 and E3 who were associated with some large effect sizes in this regard.

The only consistent trend for the two experimenters for whom data were available (see Table 7), suggested, with medium and large effect sizes (r = - .325, - .517), that more relaxed experimenters were associated with higher target ratings. More optimistic and more confident experimenter ratings were also associated with higher target ratings, though the effect sizes were generally very small (r = .075 to .223).

Perceived rapport and quality of interaction

Again there were no consistent findings across all three experimenters for the receivers' ratings (see Table 5). There was some indication that for E1, receivers' ratings of more positive (r = .346) and more spontaneous (r = .425) interaction between the experimenter and the participants was associated with higher target ratings, but for E2 and/or E3 poorer rapport with the experimenter or other participant or colder or more rehearsed interactions were associated with higher target ratings.

Although there were some inconsistencies across the experimenters, the general overall trend for senders' ratings (see Table 6) suggested that greater perceived rapport and warmer, more spontaneous and more positive interactions between the experimenter and the participants were associated with higher target ratings relative to the dummy clips, with small to medium effect sizes in most cases.

For both experimenters for whom data were available (see Table 7), the trend suggested that perceptions of warmer, more spontaneous and more positive interactions with the participants were associated with higher target ratings, although the effect sizes were generally small (range .038 to .406). It is interesting to note that for E1 there was a large positive correlation between quality of rapport with the other participants and target ratings (r = .595) but almost no similar relationship for E3 (r = - .017).

DISCUSSION

Overall, as with our previous ganzfeld study, there was no overall evidence for ESP and receivers did not give higher ratings to the target relative to the ratings given to the dummy video-clips; in fact such ratings were slightly lower. There were negligible differences between the experimenters' results with NH's and CR's results being almost exactly at chance, albeit very slightly above, and SS's being below chance expectations.

The findings concerning potential links between participants' and experimenters' ratings of their pre-session personal feelings and expectations and their perceptions of the rapport and quality of interaction between all parties are somewhat unclear and in some cases, even contradictory. Perhaps this is due to particular idiosyncrasies on the part of the individual experimenters in terms of how they tend to behave with participants or interactions between the moods/personalities of all those involved. Perhaps this is not surprising given that there is no evidence for psi in this experiment either overall or for any individual experimenter. Our feeling is that such a quantitative approach to assessing the link between experimenter and participant interactions has been rather unsuccessful and perhaps it is, by definition, unable to adequately tackle the complexity and quality and richness of human verbal and non-verbal interactions. The complexity of potential relationships and interactions among the variables of interest could perhaps be investigated by more multivariate analyses – this is something that we plan to look into. One of the problems with this kind of research "is that what seems cold and off-putting to one person may seem agreeably warm to another." (Schmeidler, 1997, p. 85). We are hopeful that a qualitative analysis of the videotapes of the pre-session interactions between our participants and ourselves as experimenters will be more fruitful and more useful in terms of identifying levels of rapport and identifying good and bad practice in terms of putting participants at ease and encouraging positive expectations of success in the subsequent experimental session. We support the endeavours of other researchers who have taken or recommended such a more qualitative approach (e.g., Schlitz, 1987; Smith, 2003a; Watt et al., 2002; White, 1997). Ideally, what are needed are more qualitative analyses of filmed interactions between clearly psi-inhibitory and psi-conducive experimenters (cf. Wiseman & Schlitz, 1997, 1999).

Of all of the roles involved in this experiment, if there is any possible trend it is for the senders' ratings suggesting that trials were slightly more successful the more relaxed, optimistic, and confident of success they were and the greater the perceived rapport and quality of interaction with the experimenter and the receiver. Particularly in the ganzfeld context, the receiver might appear to have the more important role given that more preparation is required for them and therefore more time invested in them. Perhaps it is important to ensure that equal (more?) attention is paid to the senders during the pre-session briefings, even though it seems possible for ESP to occur without the presence of a sender. Perhaps the receiver is more sensitive to the experimenter and this is why their results are more variable. There is also a suggestion ($r = -.332$, $p = .113$, two-tailed), albeit a non-significant one, that the more relaxed the experimenters feel before the session the higher the ratings given to the target clip at the end of the sessions. One aspect that both CR and SS found difficult, at times, was focusing fully on sessions that were scheduled during busy teaching and administration days when other tasks were also on their minds; some sessions were scheduled in the evenings and on weekends, but the scheduling was mostly determined by the availability of the participants. This seems to provide further support for more attention being paid to pre-session preparation by the experimenters (e.g., Delanoy, 1997; Schlitz, 1987; Watt et al., 2002).

Based upon the participants' and experimenters' ratings, it is clear that all those involved tended to be fairly relaxed and in a positive mood at the end of the pre-session briefing and perceived the quality of the interactions to have been warm, fairly spontaneous and positive. It is possible that NH might have had a better rapport with some participants because of her contact with participants prior to the session; NH was responsible for arranging the majority of the sessions. Neither the participants nor the experimenters rated their confidence that the session would be a success as being extremely high, though the experimenters' ratings tended to be higher than the participants'. Although there was a warm social ambience here, it would seem that high expectations of success at this particular task were lacking as far as the participants were concerned (and to some extent from the experimenters--see Dalton's, 1997, recommendations). Perhaps extremely highly-motivated experimenters with extremely high expectations of success who do not question the existence of psi at all, and who have readily integrated it into their daily lives, are required (e. g., see Schmeidler, 1987; White, 1976). It may well be a case of pre-selecting experimenters as well as participants with 'the right stuff.'

ACKNOWLEDGEMENT

We would like to thank the Perrott-Warrick Fund for their support of this project.

REFERENCES

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*, 4-18.

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report*. Hove, East Sussex: Psychology Press.

Dalton, K. (1997). Is there a formula to success in the ganzfeld? Observations on predictors of psi-ganzfeld performance. *European Journal of Parapsychology*, 13, 71-82.

Delanoy, D. L. (1997). Important psi-conducive practices and issues: Impressions from six parapsychological laboratories. *European Journal of Parapsychology*, 13, 63-70.

Edge, H., & Farkash, M. (1982). Further support for the psi-distributed hypothesis [Abstract]. In W. G. Roll, R. L. Morris, & R. A. White (Eds.), *Research in parapsychology 1981* (pp. 171-172). Metuchen, NJ: Scarecrow Press.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the Psychoticism scale. *Personality and Individual Differences*, 6, 21-29.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, 97, 363-386.

Honorton, C., Ramsey, M., & Cabibbo, C. (1975). Experimenter effects in extrasensory perception. *Journal of the American Society for Psychical Research*, 69, 135-149.

Honorton, C., Berger, R., Varvoglis, M., Quant, M., Derr, P., Schecter, E., & Ferrari, D. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology*, 54, 99-139.

Irwin, H. J. (1999). Experimental research on extrasensory perception. In, *An introduction to parapsychology* (3rd ed.) (pp. 59-111). Jefferson, NC: McFarland.

Kennedy, J. E. & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, 40, 1-33.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, 125, 387-391.

Morris. R. L., Dalton, K., Delanoy, D.L., & Watt, C. (1995). Comparison of the sender / no sender condition in the ganzfeld. *Proceedings of the Parapsychological Association 38th Annual Convention*, 244-259.

Murphy, G. I. (1949). Psychical research and human personality. *Proceedings of the Society for Psychical Research*, 48, 1-15.

Palmer, J. (1986). ESP research findings: The process approach. In Edge, H. L., Morris, R. L., Palmer, J. & Rush, J. H. (Eds.) *Foundations of parapsychology* (pp. 184-222). London: Routledge & Kegan Paul.

Palmer, J. (1997a). The challenge of experimenter psi. *European Journal of Parapsychology*, 13, 110-125.

Palmer, J. (1997b). Correlates of ESP magnitude and direction in the PRL and RRC autoganzfeld data bases *Proceedings of the Parapsychological Association 40th Annual Convention*, 283-298.

Palmer, J. (unpub.). *The psychology of ESP: Magnitude vs direction*. Unpublished manuscript.

Parker, A. (1975). A pilot study of the influence of experimenter expectancy on ESP scores. In J.D. Morris, W.G. Roll, & R.L. Morris (Eds), *Research in Parapsychology 1974*, (pp 42-44). Metuchen, NJ: Scarecrow Press.

Parker, A. (1977). Parapsychologists' personality and psi in relation to the experimenter effect [Abstract]. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 107-109). Metuchen, NJ: Scarecrow Press.

Parker, A., Millar, B., & Beloff, J. (1977). A three-experimenter ganzfeld: An attempt to use the ganzfeld technique to study the experimenter effect. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 52-54). Metuchen, NJ: Scarecrow Press.

Pratt, J. G. & Price, M. M. (1938). The experimenter-subject relationship in tests for ESP. *Journal of Parapsychology*, 2, 84-94.

Rhine, J. B. (1948). Conditions favoring success in psi tests. *Journal of Parapsychology*, 12, 58-75.

Roe, C. A., & Sherwood, S. J., & Holt, N. J. (2003). Interpersonal psi: Exploring the role of the sender in Ganzfeld GESP tasks. *Proceedings of the Parapsychological Association 46th Annual Convention*, 237-256.

Rosenthal, R. (1966). *Experimenter effects in behavioural research.* New York: Appleton-Century-Crofts.

Schlitz, M. (1987). An ethnographic approach to the study of psi: Methodology and preliminary data [Abstract]. In D. H. Weiner, & R. D. Nelson (Eds.), *Research in parapsychology 1986* (pp. 103-106). Metuchen, NJ: Scarecrow Press.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and mismatches.* Jefferson, NC: McFarland.

Schmeidler, G. R. (1997). Psi-conducive experimenters and psi-permissive ones. *European Journal of Parapsychology*, 13, 83-94.

Schmeidler, G. R. & Maher, M. (1981). Judges responses to the nonverbal behaviour of psi-conducive and psi-inhibitory experimenters. *Journal of the American Society for Psychical Research*, 745, 241-257.

Schneider, R., Binder, M., & Walach, H. (2000). Examining the role of neutral versus personal experimenter-participant interactions: A EDA-DMILS experiment. *Journal of Parapsychology*, 64, 181-194.

Sharp, V. & Clark, C. C. (1937). Group tests for extrasensory perception. *Journal of Parapsychology*, 1, 123-142.

Smith, M. D. (2003a). The role of the experimenter in parapsychological research. *Journal of Consciousness Studies*, 10, 69-84.

Smith, M. D. (2003b). The psychology of the “psi-conducive“ experimenter: Personality, attitudes towards psi, and personal psi experience. *Journal of Parapsychology*, 67, 117-128.

Taddonio, J. L. (1976). The relationship of experimenter expectancy to performance on ESP tasks. *Journal of Parapsychology*, 40, 107-114.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research*, 59, 172-186.

Watt, C. & Baker, I. S. (2002). Remote facilitation of attention focusing with psi-supportive versus psi-unsupportive experimenter suggestions. *Journal of Parapsychology*, 66, 151-168.

Watt, C., Wiseman, R., & Schlitz, M. (2002). Tacit information in remote staring research: The Wiseman-Schlitz interviews. *The Paranormal Review*, 24, 18-25.

White, R. A. (1976). The limits of experimenter influence on psi test results: Can any be set? *Journal of the American Society for Psychical Research*, 70, 333-369

White, R. A. (1977). The influence of the experimenter motivation, attitudes and methods of handling subjects in psi test results (pp. 273-304). In B.B. Wolman (Ed.), *Handbook of parapsychology*. New York: Van Nostrand Reinhold.

White, R. A. (1997). Suggestions for exploring and recording the inscape of psi researchers. *European Journal of Parapsychology*, 13, 96-109.

Wiseman, R. & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology*, 61, 197-207.

Wiseman, R. & Schlitz, M. (1999). Experimenter effects and the remote detection of staring: An attempted replication. In the *Proceedings of Presented Papers of the 42nd Parapsychological Association 42nd Annual Convention*, 471-279.

Woodruff, J. L. & Dale, L. A. (1950). Subject and experimenter attitudes in relation to ESP scoring. *Journal of the American Society for Psychical Research*, 44, 87-112.

Address for correspondence: Simon J. Sherwood,
Division of Psychology, University College
Northampton, Park Campus, Boughton Green Road,
Northampton NN2 7AL, UK.
E-mail: simon.sherwood@northampton.ac.uk

DO YOU KNOW WHO IS ON THE PHONE? REPLICATION OF AN EXPERIMENT ON TELEPHONE TELEPATHY

Stefan Schmidt, Susanne Müller, & Harald Walach
Department of Evaluation Research in Complementary Medicine
University Hospital Freiburg

ABSTRACT

Many people report that they know in advance who is on the phone when the telephone is ringing. Such reports may be explained by selective memory or expectancy effects but there are also examples that resist such hypotheses. Thus many people believe in an extrasensory communication transfer that may be termed 'telephone telepathy'. Surveys show that this kind of belief is widespread and might be one of the most common beliefs regarding the paranormal.

Rupert Sheldrake conducted several experiments to find out whether this effect is really due to ESP. Subjects had to determine which one of four possible callers is on the phone while the telephone was still ringing. Sheldrake reports highly significant hit rates that cannot be explained by conventional theories. He claims furthermore that callers who are familiar to the person answering the phone are identified at significant better rates than unfamiliar callers.

We attempted to replicate both of these findings by setting up a replication experiment. Twenty-one participants were twice invited for a two-hour session in an office-like room and were asked to pre-identify the callers of 10 phone calls during each session. The caller could be either one of two persons known to the participant or one of two persons unknown to them. With these four possible callers there is a mean chance expectation of 25% correct guesses. Participants were asked to complete questionnaires on earlier experiences with 'telephone telepathy', mood, personality and paranormal belief. Then participants spent approximately 100 minutes in the room together with an experimenter while the whole session was recorded on videotape. Approximately every ten minutes a telephone rang and the participant had to announce his call before the experimenter lifted the receiver to check who was calling. The sequence of the four possible callers was determined by a random event generator.

Overall valid responses were obtained on 397 calls and the participants identified 106 calls (26.7%) correctly. This result is not significant ($z = 0.78$). However, 67 (63.2 %) of the correct calls were by familiar callers showing that this group was identified more often than the unknown callers. But this result can be explained by a response bias in the participants because in all 397 calls they responded 242 times (61.0%) with the name of a familiar person. This resulted in a non-significant hit rate of 27.7% and also in a non-significant hit rate of 25.2 % ($z = 0.05$, n.s.) for the unfamiliar callers respectively. The difference between these two hit rates itself is not significant. Thus, this replication attempt failed to yield any telepathic effect. The possible reasons for the non-significant results are discussed.

INTRODUCTION

Many people report strange experiences when called by or calling others on the telephone. There are reports of situations where a person intends to call another person and in the very moment when s/he wants to lift the receiver this other person calls in. Or the telephone rings and the called person has an intuitive knowledge who is calling when s/he answers the phone. Or somebody is accidentally thinking about an old friend from school s/he hasn't heard from some twenty years just to find a message of this person on the answering machine the same evening. Friends call each other in exactly the same moment and both find the line engaged. Some of these experiences seem to be beyond any reasonable explanations, such as expected phone calls or regular calling schedules and leave the affected persons stunned.

Rupert Sheldrake and colleagues conducted two surveys on this phenomena in California and the UK (Brown & Sheldrake, 2001; Sheldrake, 2000). Amongst other questions they asked both populations: "Have you ever heard the telephone ring or picked up the telephone and known who was on the other end without possible cue, before they have spoken?" In Lancashire, 49% of 200 persons asked answered "yes" while in California 47% answered "yes". While some of these experiences might be explained by coincidences, many people tend to believe in unconventional explanations because their experiences were either too extraordinary to attribute it to a coincidence or just made sense in a perfect way. Thus Sheldrake calls these phenomena "... the commonest kind of apparent telepathy in the modern world." (Sheldrake & Smart, 2003a, p. 184). He was also the first person to conduct an experiment on this phenomenon.

In this simple experiment (Sheldrake & Smart, 2003a) a subject gets called by one of four pre-selected callers. The callers are selected by a random process and there is no conventional possibility to know in advance whom the caller will be. The subject guesses the name of the caller before answering the phone. Within this simple experiment there is a mean chance expectation (MCE) of $p = .25$ to make a correct guess. If the above reported phenomena are just due to chance, unconscious expectancies, selective memory of these events or other conventional explanations then exactly this mean chance expectation should be found if the experiment is conducted thoroughly and there is no sensory leakage of the relevant information. But if on the other hand the intuitive knowledge on the side of the callee is due to some unknown information transmission process (that might be called telepathy or just Psi), then the MCE should be exceeded in such an experiment, provided that the phenomena shows up on demand.

It is the merit of Rupert Sheldrake to develop such a simple and clear-cut design and to conduct this straightforward test for "telephone telepathy" as he calls it. In his first publication, he and his colleague report on several series of altogether 571 trials with 63 different participants (Sheldrake & Smart, 2003a). Four different experimental protocols were applied in these series, but the basic task to identify one of four potential callers was maintained. Overall there were 231 correct guesses (hit rate 40 %, $p = 4 \times 10^{-16}$) [1]. Thus the telepathy hypothesis was apparently confirmed by an extraordinarily strong effect. However, on close examination this study showed some limitations. One weakness was that the experiment was rather uncontrolled. Participants waited at home for about an hour for the target person to call. After the call the experimenter phoned the participant and asked what s/he had guessed. In some trials this was counterchecked with the respective caller but, nevertheless, this procedure allows for a whole range of manipulations on the side of the participant. A second limitation is more serious. When the telephone rang the participants lifted first the receiver and then spoke aloud the name of the guessed person into the receiver. This procedure of defining the guess after lifting the receiver allows for sensory leakage. A typical clicking, a humming sound or some background noise from the caller might give a relevant clue.

But a second study (Sheldrake & Smart, 2003b) with four participants was controlled much tighter. For this study only selected participants were employed. Every participant had to reach a minimum success rate in a pretest with ten trials. Four participants remained for a total of 271 trials (183 trials by only one participant). In these trials the subjects were videotaped while awaiting the call. They spoke their guess into the video camera before lifting the receiver and thus the data of this study can be considered as a valid and good controlled examination of the hypothesis on telephone telepathy. Overall there were 122 correct guesses (hit rate 45%, $p = 10^{-12}$) and the authors conclude this to be a proof for telepathy. They also tested a second hypothesis on the difference between callers who are known to the participants and others who are not. Therefore, in 175 trials only two of the four callers were familiar to the participants while the other two were recruited by the experimenter. Sheldrake and Smart hypothesize that only towards the familiar callers there will be a telepathic bond with an increased hit rate, while the unfamiliar callers will be identified no better than chance. They report hit rates of 61% for familiar callers and 20 % for unfamiliar callers with a significant difference ($p = 8 \times 10^{-8}$). However the statistical procedure they applied proved to be unsatisfactory. The data were not corrected for a response bias. Upon inspection of the data it could be shown that the participants indeed guessed more often familiar callers than unfamiliar ones and this resulted in a higher probability to obtain a hit with a familiar caller (Schmidt, Müller & Walach, 2003). When corrected for this bias the hit rates resulted in 46% hit rate for familiar callers and 37% for unfamiliar callers respectively. The difference between these hit rates was no longer significant ($p = .32$), if

[1] A recalculation by us yielded a slightly smaller but nevertheless highly significant p-value of 1.7×10^{-14}

the same statistical test that was also used by Sheldrake & Smart (2003b) was applied. However, in a reply to this critique Rupert Sheldrake showed that the data are reaching significance again if a different statistical analysis is applied (randomization test, p=0.0002). With the data being significant in one analysis and not significant in the other, the question whether familiar callers are identified better than unfamiliar ones remains open.

Altogether the Sheldrake studies have a rather exploratory nature with frequent changes in the design and only partially prespecified number of trials. We thus set out to conduct a more rigorous conceptual replication of these two findings (telephone telepathy effect, better results with familiar callers). In this experiment we repeated the basic design of the Sheldrake studies. We invited participants who report to be familiar with the phenomenon and asked them to name two callers. Two other callers were assigned by the experimenter and thus the participant had the task to identify the correct caller out of a possible four of which only two were familiar to him or her. However, to improve on experimental controls and to make the data collection more efficient we also applied changes to the original protocol. Instead of visiting the participants at home we invited them to an office. This provided us with the possibility to conduct all trials under the same controlled conditions, but on the other hand also reduced the ecological validity of the study. We furthermore carried out ten consecutive trials at once within a two-hour session. We also decided to work with a self-selected sample rather than to identify high-scorers by pre-tests. This approach was chosen to guarantee that the results generalize to the reports from everyday life. Another change applied to the role of the experimenter. In contrast to the Sheldrake studies, where there was no experimenter present and where the participants answered the call, the experimenter in our study -stayed with the participant all the time during the experiment. The experimenter also lifted the receiver for the incoming calls after the participant's guess was recorded to find out who was calling. We did so to avoid a sensory leakage in a possible participant-caller interaction. This was necessary because callers knew whether they had to call again in the same session or not and thus might drop accidentally verbal cues such as "till later" or "OK. We are done!"

METHODS

Design and Hypotheses

We planned to conduct a study with a preplanned number of 400 trials and 20 participants on the basis of a power analysis of the Sheldrake studies. Participants were invited for two sessions of 10 trials each. Hypothesis 1 stated that participants are able to identify the correct caller out of a possible four significantly better than expected by chance (MCE = 25%). Hypothesis 2 stated that participants are able to identify familiar callers significantly better than unfamiliar ones.

Participants

Twenty-one participants were recruited through newspaper advertisement, press reports and leaflets. Nineteen participants contributed 20 trials in two sessions of ten. Two participants took part in only one session of ten trials. Seventeen participants (81%) were female and the mean age of all participants was 39.8 years (SD = 11.3, range 22-62).

Experimenters

Three different experimenters participated in the study. One of them, a 35 year-old psychologist served as the experimenter for the interaction with the participants for all sessions. The other two (the second author SM and a student) were in charge of scheduling the callers during the session and had no direct interaction with the participants.

The complete organization and conductance of this study was carried out by the second author, SM (Müller, 2003).

Facilities

The experiment took place in an office space of a youth counselling organisation in Freiburg, Germany. The office contained a video camera for the monitoring of the session. The telephone for the incoming calls was placed on a table. The telephone had no display for caller identification. The room was further equipped with reading material for the participants and some cookies.

For the randomization and scheduling of the callers a second office was used in the University Hospital Freiburg. This office was located in a distance of approximately 3 km from the first one.

Material

Participants were asked to complete the following questionnaires:

1. A questionnaire similar to the one presented by Sheldrake (2000) asking about prior experiences with 'telephone telepathy'. In addition we added some questions on the expectancy of success in the upcoming experiment.
2. A German sheep-goat scale. We applied the questionnaire on paranormal beliefs by Brednich (1993), that was already applied in other German Psi studies (Schmidt, Schneider, Binder, Bürkle & Walach, 2001; Schmidt, Tippenhauer & Walach, 2001; Walach & Schmidt, 1996). This is 21-item scale with various statements on ESP, reincarnation, magic and astrology that have to be rated on a four-point scale.
3. The German translation of the NEO-FFI (Borkenau & Ostendorf, 1993) a sixty-item personality inventory assessing the five standard traits: neuroticism, extraversion, openness, conscientiousness and agreeableness.
4. The Basler Mood Scale/Basler Befindlichkeitsskala, (Hobi, 1985). A 16-item scale assessing the present mood of the participant on the basis of a subjective rating. The scale has four factors: vitality, balance, alertness and social extraversion.

Procedures

Participants interested in the advertised study called the experimenter SM and were asked whether they already had experiences of knowing who was on the phone in the absence of sensory cues and whether they would be able to provide two callers to participate in the experiment. If these criteria were fulfilled, two sessions were scheduled. Upon arrival at the location the participants were greeted by the experimenter and informed about the procedures of the session. Then the participant was asked to fill in the questionnaires. The first three questionnaires were only filled in once during the first visit. The Basler Mood Scale was filled in prior to both experimental trials.

While the participant filled in the questionnaires, the experimenter staying with the participant switched on the video camera and then phoned the second experimenter in the second office at the hospital. The second experimenter then collected a sealed envelope from a locked cabinet drawer that contained the randomized calling sequence for the session. The sequence consisted of the numbers 1, 2, 3, and 4, and were assigned as follows to the four callers: 1 to the first familiar caller that was nominated by the subject, 2 to the second familiar caller, 3 to the first unfamiliar caller chosen by the experimenter and 4 to the second one. The ten calls were then scheduled 10 minutes apart. The scheduling experimenter served as one of the unfamiliar callers. The other three callers were then called by the experimenter and informed about the telephone number to call and the calling time. If a caller was chosen by the randomization to conduct more than one call then the schedule for all calls was given at once. Once the scheduling was completed, the scheduling experimenter phoned back to the experimenter staying with the participant and the experiment was ready to start. During the next 90 minutes the participant received ten calls approximately one every ten minutes. When the telephone rang the participant spoke his or her guess into the video camera and the experimenter recorded the guess on a session sheet. Then the experimenter lifted the receiver and checked for the caller. This information was fed back to the participant. Between the calls the participants were free to read some of the magazines provided, whose content was unrelated to the experiment, or they could have a chat with the experimenter.

When the participant arrived for the second sessions the procedures were the same except that the participants this time only had to fill in the mood scale but not the other questionnaires.

Randomization, Data analysis and Statistics

For the randomization, the Marsaglia-Zaman random number generator was employed (Marsaglia & Zaman, 1987). This RNG consists of a software algorithm that deduces two seeds from the computer clock and produces random sequences that have proven to be perfectly at chance. Fifty sequences of ten number sequences consisting of the numbers from one to four were generated before the first session of the experiment and placed in opaque envelopes labeled with a number from 1-50. These envelopes were stored in a locked cabinet drawer. The experimenter always chosed the envelope labeled with the smallest number.

Questionnaire data, calling sequences and guesses of the participants were entered into SPSS for Windows 11.0. Further analyses were conducted in SPSS and in Excel.

Statistics: The mean chance expectation for each single trial is $p = .25$ and within each trial all four targets have the same probability independently of the prior trials (open deck). For Hypothesis 1, the likelihood of X hits within N trials can be calculated by the classical formula for the critical ratio (*CR*): $CR = (X - Np)/sqrt(Npq)$ (with $q=1-p$) (Burdick & Kelly, 1977). *CR* is equivalent to a *z*-score and normally distributed. The *CR* can be calculated for any given number of trials independently whether they are from one or several participants. This analysis was the prespecified method for the evaluation of hypothesis 1. The test of significance was one-tailed. Another possibility to assess the results is the application of a binominal test calculating the likelihood to obtain X or lesser hits out of N trials, when the probability to obtain a hit is at $p = .25$.

For Hypothesis 2 we calculated a χ^2-test for the distribution of the data in 2 x 2 table with the rows hits (yes – no) and the columns calls (familiar – unfamiliar). This test returns a *p*-value for the likelihood that the distribution of the hits is independent from the expectancy for a familiar or unfamiliar caller.

RESULTS

We obtained 397 valid trials. Data from three trials were excluded from the analysis because the call arrived too late. The calls came from the following callers: familiar caller 1 (f1): 103 calls; familiar caller 2 (f2): 98 calls; unfamiliar caller 1 (u1): 101 calls; unfamiliar caller 2 (u2): 95. These numbers prove that the random distribution was indeed at random.

On the other hand, the distribution of the guesses for these four callers by the participants was not at random. Participants guessed in most cases the caller to be a familiar one with an additional preference for the first familiar caller they had named. The guesses were: guessed for f1: 131; guesses for f2: 111, guesses for u1: 96; guesses for u2: 59. This distribution shows a response bias.

Hypothesis 1: The participants identified 106 of the 397 calls correctly. This results in a hit rate of 26.7 % (MCE = 25%) and a critical ratio of $z = .78$. The according p-value (one tailed) is $p = .22$. The binominal test arrives at a similar p-value of $p = .23$. Thus, hypothesis 1 was not confirmed.

Out of 19 participants that completed all 20 trials two arrived at an individual significant hit rate of 9 and 10 hits respectively. The two participants completing only ten trials had a non-significant score of two hits each.

Hypothesis 2: From the 106 correct guesses 67 (63.2%) were with familiar callers and 39 (36.8%) with unfamiliar callers. If based on the distribution of the incoming calls this would result in a hit rate of 33.3 % for familiar callers and 19.9 % for unfamiliar caller. In other words when a familiar caller called s/he was identified in 33.3% of all trials correctly as familiar. But it was already said in the introduction that such a procedure to base the hit rates on the distribution of the incoming calls is prone to error because it does not take into account the response bias of the participants. If the hit rates are based on the participants' guesses for familiar or unfamiliar callers the hit rates change to the following values: familiar callers: 27.7 % (n.s.) and unfamiliar callers: 25.2 % (n.s.). The according 2 x2 table looks as follows (see Table 1).

		Hit		
		Yes	No	Sum
Guesses	Familiar	67	175	242
	Unfamiliar	39	116	155
	Sum	106	291	397

Table 1: Crosstabulation of hits and guesses for familiar and unfamiliar callers.

The χ^2-test for this table results in a non-significant p-value of p = .58 (χ^2 = 0.31, df =1) and thus hypothesis two was not confirmed.

For the sake of completeness we provide the raw data of all 21 subjects in Table 2.

Part. No.	Trials	Guess Fam.	Guess Unfam.	Hits Fam.	Hits Unfam.	Total Hits	Total Hits %
1	20	10	10	2	3	5	20
2	19	11	8	3	2	5	26
3	19	15	4	3	2	5	26
4	19	12	7	4	2	6	32
5	20	10	10	3	0	3	15
6	10	6	4	1	1	2	20
7	20	8	12	1	3	4	20
8	20	13	7	6	2	8	40
9	10	7	3	1	1	2	20
10	20	12	8	4	0	4	20
11	20	12	8	4	3	7	35
12	20	14	6	5	1	6	30
13	20	14	6	2	1	3	15
14	20	13	7	1	0	1	05
15	20	11	9	3	1	4	20
16	20	14	6	3	2	5	25
17	20	13	7	6	3	9	45
18	20	13	7	7	3	10	50
19	20	16	4	2	1	3	15
20	20	8	12	4	4	8	40
21	20	10	10	2	4	6	30
total	**397**	**242**	**155**	**67**	**39**	**106**	

Table 2: Complete data for all 21 subjects and 397 valid trials. guess fam. = number of guesses for familiar callers; guess unfam. = number of guesses for unfamiliar callers; hits fam. = number of hits for incoming calls of familiar callers; hits unfam. = number of hits for incoming calls of unfamiliar callers.

DISCUSSION

Our task to replicate the positive findings on telephone telepathy reported by Rupert Sheldrake in a tighter controlled setting failed. There was neither a hit rate larger than expected under chance condition, nor did we find any significant difference between the hit rates for familiar and unfamiliar callers. The data obtained in our study match almost perfectly chance expectation and do not allow for any interpretation of unconventional information transfer.

This result is in accordance with the hypothesis that all accounts for strange experiences as described in the introduction are due to coincidences, selective memory or unconscious expectancy and the like. But this result is definitive not in accordance with the findings reported by Sheldrake and Smart (2003a, 2003b).

There are various interpretations to this inconsistency. One would be that there is no such effect like telephone telepathy and that the data presented by Sheldrake are due to some methodological error. The second one would be that one of the changes to the design that were made by us in our undertaking to establish a conceptual replication was fatal and avoided the effect demonstrated by Sheldrake to show up in our study. A third interpretation would centre on the replication problem within parapsychology and would speculate about basic properties of Psi that avoid replication beyond ones that can be attributed to changes in the design.

The first interpretation that the Sheldrake findings are based on erroneous methods is of course one possibility. But one has to acknowledge that the experimental procedures described by Sheldrake and Smart (2003a, 2003b) show a steady increase in controls and thoroughness from the first pilot trials to the final design with videotaped sessions and significant effects are reported for all of these different protocols. Of course one can never be sure what really happens in these studies until one witnesses the data collection, but from the information that was available to us these data look valid and cannot be explained away.

Thus, the second interpretation stating that differences in the design between our study and the ones by Sheldrake are responsible for our failure to replicate Sheldrake's result is becoming more likely. The main changes in our experiment were: (i) The experiment took place in a room provided by us rather than in the participants' homes. (ii) The participants received ten calls and not one per session. (iii) The experimenter was present during the session and also lifted the receiver instead of the participants. (iv) Selection of sample: we used a self-selected sample rather than to restrict the experiment to high scorers who have passed a pretest.

The first three changes reduced the ecological validity of the study (in favour for better controls) and thus one could conclude that the phenomenon shows up under natural conditions (participants at home by themselves getting one phone call) but not under the more artificial conditions in the setting we provided (participant receiving phone calls every ten minutes in an office that is alien to them). On the other hand, one could argue that Sheldrake also obtained significant findings in sessions that were placed and broadcasted from a TV studio. In Sheldrake's earlier trials participants also were called twice and the performance did not decrease (although one has to mention that there is some likelihood for sensory cueing in these trials).

Whether the differences introduced by the changes (i) and (iii) are really responsible for our failure to replicate can be only checked through a direct comparison of these conditions in a follow-up study. But for the change in the number of successive calls (ii), we can check whether participants were more successful in their first trials. Figure 1 shows the hit rate per trial (bars) and also a curve representing the cumulative hit rates for the trials 1 to 10 of a session. From the cumulative curve one can extract the results if one would stop after a shorter number of trials than 10 in one session. For only one trial would one obtain the highest hit rate (30.8%, p = .25). If our participants had shown the same hit rate in all 397 trials as in their first trials, we would arrive at a significant p-value of p = .006. Thus it may also be that the difference between our results and the results by Rupert Sheldrake are at least partially due to the fact that participants performance decreased after their first trial. One reasonable explanation would be that participants get tired over time. On the other hand, one can see that this might also be due to a normal fluctuation. An inspection of the average hit rate in the different trials (bars) demonstrates that participants performed even better in their 6th and also in their 9th trial than in their first one.

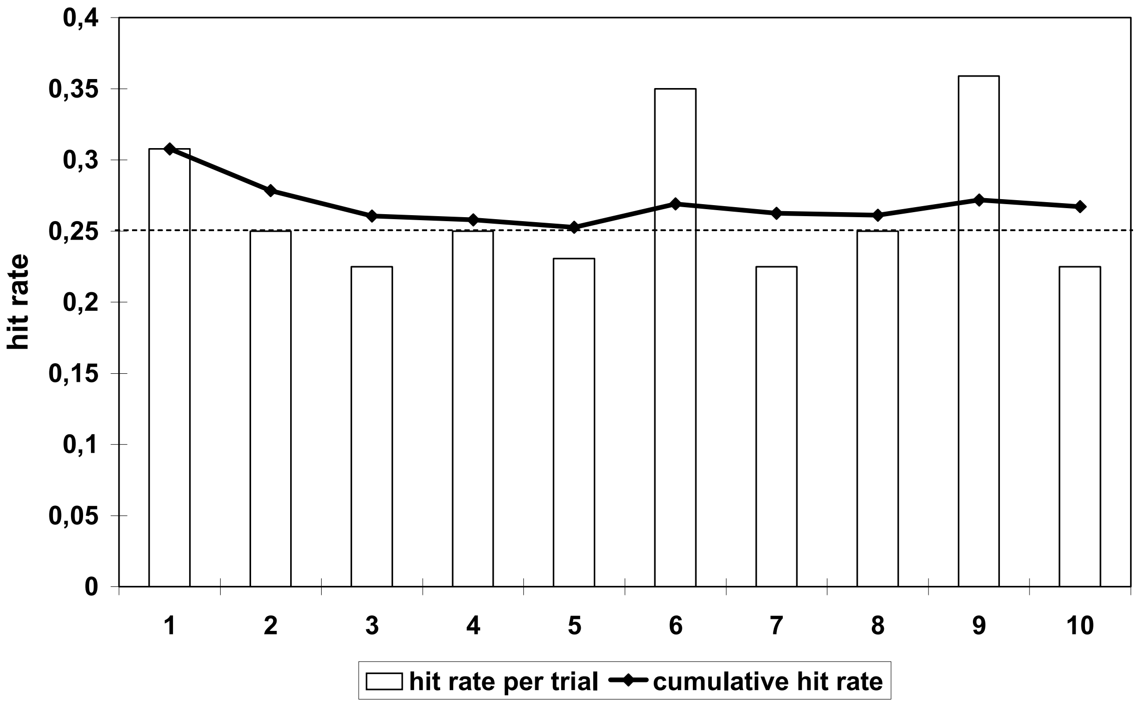


Figure 1: Hit rate per trial (bars) and cumulative hit rate (curve) over the successive 10 trials in a session. The dotted line indicates the MCE of .25.

The change in location and the presence of the experimenter also might have been changes that mattered. One of the safeguards in our study was that the experimenter, rather than the participant (and thus receiver of the call), lifted the receiver and talked to the callers. One could argue that by this procedure the effect was avoided because the caller did not have the chance anymore to reach the participant. But on the other side one could argue as well that the meaning context of the experiment was still intact. The participant sat in front of the telephone and tried to find out who is responsible for the ringing of the phone. The change in the procedure occurred after the recording of the guess and furthermore everybody knew about this circumstance. The point in this discussion is that to find out which of these two lines of reasoning is the right one, one has to know the precise mechanism of the alleged effect. Until this mechanism is known one can not give a precise experimental description how to find the effect and also does not know what effect size under which conditions can be expected. Collins calls this paradox the experimenter regress (Collins, 1985).

An important difference was also the selection of participants (iv). One could argue that Sheldrake has selected sensitive participants through his process of performing a pre-test. Other researchers have already reported that telepathy is not necessarily normally distributed in the population but restricted to a few gifted subjects (see e.g. Utts, 1996). Our self-selected sample probably did not contain any high scoring participant (or maybe not enough). But if this were true then the many reports of strange experiences around receiving phone calls could not anymore be attributed to ‘telephone telepathy’. If this is really “... the commonest kind of apparent telepathy in the modern world” that is reported by approx. 50% of the population (see Introduction) then the experiment should also work with a self-selected sample that has experience with this phenomenon. It was our specific aim to study this everyday life situation and we recruited our sample accordingly.

Finally there remains the possibility that our result is due to an intrinsic property of Psi not to show up stable within the same experimental context. Not to find the same effect again in a replication is a well-known result within parapsychology and has also occurred in our work before (see e.g. Schmidt, Tippenhauer & Walach, 2001). A theoretical framework for this can be found, for example, *within Lucadou's Model of Pragmatic Information (Lucadou, 1995, 2001). Whether such a circumstance is responsible for our results can be only determined by the development and empirical testing of a model for Psi that accounts for such intrinsic properties.

CONCLUSION

The positive findings by Sheldrake and Smart (2003a, b) could not be replicated in our experiment. This may be due to changes that were made to the design due to our task to conduct a conceptual replication with tighter controls. Whether these changes were responsible for our failure to replicate earlier findings cannot be clearly decided. This can only be determined by a replication where these conditions can be compared (e.g. the results of the same participants are compared in the experimenter's office and in their home, or the results of the same participants are compared in a condition with an experimenter present or not present). We suggest conducting further studies under various conditions by different experimenters and researchers to shed light on conducive and inhibitory conditions of the alleged effect.

REFERENCES

Borkenau, P. & Ostendorf, F. (1993). *Neo-Fünf-Faktoren Inventar*. Göttingen: Hogrefe.

Brednich, A. (1993). Eine Fragebogenuntersuchung zur Erfassung von magisch-irrationalem Denken und schizotypischen Persönlichkeitsstörungen bei Erwachsenen. Unveröffentlichte Diplomarbeit, Albert-Ludwis-Universität Freiburg i. Br.

Brown, D. J., & Sheldrake, R. (2001). The anticipation of telephone calls: A survey in california. *Journal of Parapsychology, 65,* 145-156.

Burdick, D. S., & Kelly, E. F. (1977). Statistical methods in parapsychological research. In: B. B. Wolman *Handbook of parapsychology* (pp. 81-130). New York, NY: Van Nostrand Reinhold.

Collins, H. M. (1985). *Changing order*. London: Sage.

Hobi, V. (1985). Basler Befindlichkeits-Skala. Ein Self-Rating zur Verlaufsmessung der Befindlichkeit. Weinheim: Beltz Test Gesellschaft.

Lucadou, W. v. (1995). The model of pragmatic information (MPI). *European Journal of Parapsychology, 11,* 58-75.

Lucadou, W. v. (2001). Hans In luck: The currency of evidence in parapsychology. *The Journal of Parapsychology, 65,* 3-16.

Marsaglia, G., & Zaman, A. (1987). Toward a universal random number generator (No. FSU-SCRI-87-50): Florida State University.

Müller, S. (2003). *Das Telefonexperiment. Die experimentelle Umsetzung eines Alltagsphänomens*. Unpublished thesis, Albert-Ludwigs-Universtität Freiburg i. Br.

Schmidt, S., Müller, S., & Walach, H. (2003). Letter to the editor. *Journal of Parapsychology, 67,* 409-411.

Schmidt, S., Schneider, R., Binder, M., Bürkle, D., & Walach, H. (2001). Investigating methodological issues in EDA-DMILS: Results from a pilot study. *Journal of Parapsychology, 65,* 59-82.

Schmidt, S., Tippenhauer, G., & Walach, H. (2001). Connectedness between participants in an experiment on distant intention. *The Parapsychological Association 44th Annual Convention. Proceedings of Presented Papers* (pp. 285-300). Durham, NC: The Parapsychological Association.

Sheldrake, R. (2000). Telepathic telephone calls: Two surveys. *Journal of the Society for Psychical Research, 64,* 224-232.

Sheldrake, R., & Smart, P. (2003a). Experimental tests for the telephone telepathy. *Journal of the Society for Psychical Research, 67,* 184-199.

Sheldrake, R., & Smart, P. (2003b). Videotaped experiments on telephone telepathy. *Journal of Parapsychology,* 147-166.

Utts, J. M. (1996). An assessment of the evidence for psychic functioning. *Journal of Scientific Exploration, 10,* 3-39.

Walach, H., & Schmidt, S. (1996). Evidence for a non-classical experimenter effect. (Abstract). *Society for Scientific Exploration Third European Meeting.*

Address for correspondence: Stefan Schmidt, University Hospital Freiburg, Germany, Department of Evaluation Research in Complementary Medicine, Institute of Environmental Medicine and Hospital Epidemiology, Hugstetter Str. 55, D-79106 Freiburg, Germany.
E-mail: stefan.schmidt@uniklinik-freiburg.de

REMARKABLE CORRESPONDENCES BETWEEN GANZFELD MENTATION AND TARGET CONTENT – PSI OR A COGNITIVE ILLUSION?

Joakim Westerlund[1], Adrian Parker[2], Jan Dalkvist[1], Anneli Goulding[2]
[1]*Department of Psychology, Stockholm University*
[2]*Department of Psychology, Göteborg University*

ABSTRACT

Remarkable correspondences between Ganzfeld mentation and target content have been reported since the early start of Ganzfeld experiments in parapsychology. These correspondences may be due to some anomalous information transfer (e.g., telepathy) or they may be due to a cognitive illusion on the part of the person who perceives them as remarkable. The present paper presents two studies conducted in order to investigate which of these two possibilities is the more probable. Both studies were based on data collected during the first formal experiment run with a new technique called Digital Ganzfeld (Goulding, Westerlund, Parker & Wackermann, 2001). In this experiment the receiver's mentation was stored as a digital audio file that was synchronized with the target videoclip (as well as with the three decoy clips, as all clips were of exactly the same length). An external judge (JW) then played the mentation file together with the different clips in the set that was used and ranked the different clips according to similarities with the mentation. The main result was close to chance with a direct hit rate of 23% (Goulding, Westerlund, Parker & Wackermann, 2004). However, at the same time as the judge did this judging, he also tried to discern any correspondences between the mentation and any of the four clips (before knowing which one was the target) that appeared to him to be very remarkable (presented as Study 1 in this paper). A total of 20 remarkable correspondences were collected. Of these, only 6 (30%) were correspondences between the mentation and the target whereas the remaining 14 (70%) were correspondences between the mentation and decoys. This result is not significantly different from chance expectation. It was argued that this result gives some support to the cognitive illusion hypothesis, but that the study was lacking in statistical power. In order to gain stronger statistical power, a group of 11 students rated each of the 20 correspondences on scales ranging from 0-100 (presented as Study 2 in the present paper). The students did not rate the correspondences that were hits (i.e., correspondences with the target) as significantly more impressive than the correspondences that were misses (i.e., correspondences with one of the decoys). This result speaks against the psi-hypothesis, but the results should be interpreted with some caution since the students could not be said to have been 100% blind. Taken together, the results of the two studies presented here help us be more conservative about concluding that remarkable correspondences between receiver mentation and target content are due to psi when they may be due to cognitive illusions (or subjective validation).

INTRODUCTION

Since the early start of Ganzfeld experiments in parapsychology, there have been reports of remarkable correspondences between Ganzfeld mentation and target content. As an example, consider the receiver's mentation that was reported by Honorton et al. (1990):

> ...Now visual patterns more like a spider web and the color. And then like the form of the veins of a windmill... Something like a spider web again. A spider web. A pattern that instead of a spider web it looks like basket weaving... (p. 125)

The target for this receiver was a film about a spider weaving its web. As a matter of fact, the receiver said the word "spider" 46 times during the mentation (Rick Berger, personal communication, August 1997). Many parapsychologists have apparently been very impressed by such correspondences. For example, John Palmer (2003) writes:

> Because a ganzfeld session comprises only one trial, success cannot be demonstrated statistically for a single session, although sometimes the correspondence of the mentation to the target is so close in detail that success is obvious. (p 54)

Of course, most parapsychologists have also been aware of the difficulties in interpreting correspondences of this kind. However impressive they may appear, it has been impossible to know what the chances are of such correspondences occurring by random processes. So even though some of these correspondences have been published, most of them have been spread by hearsay and by informal presentations at PA conferences. One example is a videotape, compiled by Kathy Dalton, that shows some of the correspondences between receiver mentation and target content that appeared in the Sender – No sender study conducted in Edinburgh in 1994 (Morris, Dalton, Delanoy & Watt, 1995). On this tape, the target video clip is shown and at the same time the mentation of the receiver can be heard. One of the most remarkable excerpts shows a man who is running through a forest; it seems that he is being hunted (at the same time, the receiver says: "Trees. People running. Fleeing."). Suddenly, the man falls down in a deep muddy pool (at the same time, the receiver says: "Falling. Muddy."). The camera zooms in on the man's face (at the same time, the receiver says: "Blond hair. 70´s hairstyle. Curly-ish. White face." – all the utterances appear to describe exactly what is being shown on the film). The next thing that happens in the clip is that the man can no longer keep his head above the surface, so he disappears into the mud (at the same time, the receiver says: "Dead man in the water."). Unfortunately, the synchronization between the video clip and the tape with the receiver's mentation was not done automatically. Kathy Dalton (personal communication, August 1997) matched the mentation tape with the video clip according to her memory of the session (in which she was one of the experimenters).

In order to solve this uncertainty, in a series of Ganzfeld experiments conducted by Parker, Persson and Haller (2000), an additional video tape recorder was installed in the lab. Attached to the tape on this machine, the sound from the receiver was automatically recorded together with the film clip that was played from the first video tape recorder to the sender. After using this for about 100 Ganzfeld sessions, Adrian Parker was able to compile a 30-minutes long tape filled with strikingly good correspondences between the receiver's mentation and the target clip. One scene, for example, shows a woman running through a forest. At exactly the same moment when the woman falls and hits her face on the ground, the receiver says: "feels like someone is falling, hitting the face on stony ground".

Two of us (JW and JD) were very impressed with these compilations of "Greatest Hits in the Ganzfeld". We soon realized, however, that there was considerable variation in how these correspondences were perceived by different persons. Some individuals (including persons with a very skeptical and persons with a very positive attitude towards psi) were not impressed at all, while others (also including persons with a very skeptical and persons with a very positive attitude towards psi) concluded that this was as close to a "proof" of psi as you could get. Of course, we also realized that even if everyone had been extremely impressed by these correspondences, they do not constitute scientific evidence for psi. There are many ways in which this kind of subjective validation could fail, the most obvious being the inability to take into full account all instances in which there were *no* correspondences between the mentation and the target.

So, what is clearly needed is some kind of *baseline* that the correspondences between the receiver mentation and the target content can be compared to. Contrary to what many parapsychologists have formerly thought, such a baseline is in fact actually not difficult to establish. The way to do this is simply to select the most impressive correspondences while in the judgment phase of a Ganzfeld experiment. In this phase, the judge (the receiver or an independent judge) is trying to compare the mentation to the four possible film clips. What the judge should do, of course, is to note all the striking correspondences between the mentation and the content of any of the four clips. If such correspondences are signs of telepathy, they should (more often than chance) turn out to be matches between the mentation and the target clip and not between the mentation and any of the decoys.

The problem with this method is that in traditional Ganzfeld experiments, the mentation is usually only recorded on an ordinary tape recorder, without any possibility to synchronize it with the target clip during the judgment phase and without any possibility whatsoever to synchronize it with the decoy clips. Fortunately with the new Digital Ganzfeld technique, recently developed in Sweden, the mentation is automatically synchronized with both the target and the three decoys during the judgment phase. This technique has been thoroughly described elsewhere (Goulding, Westerlund, Parker & Wackermann, 2001;

Goulding, Westerlund, Parker & Wackermann, 2004; Parker, 2003), but since both studies reported in the present paper are based on data collected during the first formal Digital Ganzfeld experiment (Goulding, Westerlund, Parker & Wackermann 2004), the most important aspects of the system will be described below.

The Digital Ganzfeld

The Digital Ganzfeld program consists of two software components, one for running a Ganzfeld experiment and the other for judging a Ganzfeld experiment. When running a Ganzfeld experiment, the software first selects one set of four film clips from a library of 25 sets. All film clips are stored digitally on the computers hard drive and all clips are exactly 2 minutes and 3 seconds long (the last three seconds of each clip shows a blank screen). When the receiver is ready, the sender starts playing the clip. At exactly the same time, the computer starts recording the receiver's mentation.

The target clip plays 7 times with 0 seconds for rewinding and after this the recording of the mentation stops. Thus, a total of 14 minutes and 21 seconds of the receiver's mentation is stored on a file on the computer's hard drive. The experimental phase is not over yet, however, because now the computer once again picks one set (but not the same as was recently used) and one film clip from the set and shows this clip seven times to the sender while recording the mentation into a new file. With this procedure two sessions are squeezed into one experiment, and the data collection is thus more economical than when only one session is run.

The software for judging the experiment is primarily designed to be used by someone who is accustomed to it and who has a great deal of time available for the judging (one session usually takes 1-2 hours to judge), but it can also be used by the receiver with some help from an experimenter. A session is judged by first downloading the corresponding mentation file from the university server. The name of the file reveals which set was being used (but of course not which clip was the target). The mentation file is then opened from the judging software, together with the four film clips in the set. The mentation file can now be played together with one or two of the film clips. The receiver's mentation becomes exactly synchronized with the film clips, so any real-time correspondences that occurred during the experimental phase can be easily identified. The idea is that this should help the judge in deciding which of the four film clips was actually the target. There is also a function for ”bookmarking” correspondences. The judge can write down what the receiver said together with information about when the receiver said it and save this as a bookmark. Later, by clicking on this bookmark, the judge can again listen to the mentation while watching and hearing the corresponding film clip at the same time.

The first experiment using Digital Ganzfeld

In the first experiment with Digital Ganzfeld (Goulding, Westerlund, Parker & Wackermann, 2004), one of us (AG) was the main experimenter and another of us (JW) worked as an external judge. The experiment was run in Gothenburg, but the judging was done in Stockholm, some 400 km away. A total of 64 receivers produced 128 mentation files. JW judged all of them, but half of the receivers also made their own judgments. The main analysis was done on the number of direct hits that the external judge produced. This was almost exactly at chance level (23%). It is worth noting, however, that the receivers themselves only produced 14% direct hits, which is significantly below mean chance expectation ($p = .05$, two-tailed exact binomial test).

STUDY 1 – INTRODUCTION

The purpose of the first study was to compare the number of subjectively remarkable correspondences that could be found between the receiver mentation and the target film clip with the number of subjectively remarkable correspondences that could be found between the receiver mentation and any of the decoy clips in the current set. If correspondences of this kind are signs of telepathy, significantly more than one fourth of them should be correspondences between the mentation and the target.

One problem, which until now has not been addressed, is how long a correspondence should be allowed to be and still be judged as *one* correspondence (and not, e.g., two or three). In the present study, we decided that a correspondence should be restricted to ”a single utterance” or ”a single meaningful sequence of utterances”. The exact limits of an utterance or of a sequence of utterances were not further defined, but the reader should note some important differences between having such short segments of correspondences compared to having complete mentations.

One major problem in judging the ”unlikelihood” of obtaining similarities between the whole mentation from a Ganzfeld session and a particular film clip is the dependency issue. Suppose, e.g., that the receiver guesses that the clip the sender is watching is from a western movie. The receiver will probably say things like: ”An American Indian”, ”A man on a horse”, ”A saloon”, ”There is a quarrel”, ”Guns” and ”Shooting”. Now, if the guess actually was a correct one, there is a good chance that several, or maybe all, of these utterances will turn out to have equivalents in the clip. This is because the utterances as well as the events on the clip are not independent of each other. If a film clip shows an Indian it is also likely to show a man on a horse, and if someone thinks about a film with an Indian, that person is also likely to think about a man on a horse. By using small segments of the mentations, this dependency problem is substantially reduced. By using small segments, the task of selecting the most impressive correspondences thus becomes easier. The drawback is, of course, that a great deal of potentially important information is lost. If you only know that the receiver said ”spider” at one occasion when the target was a spider, you *should* be less impressed than if you know that the receiver said ”spider” 46 times when the target was a spider.

METHOD

While judging the session in the first Digital Ganzfeld experiment, the external judge bookmarked what appeared to him to be remarkable correspondences between the mentation and any of the four film clips in the current set. There were no limitations concerning the length or content of what should constitute one correspondence. It turned out, however, that no single correspondence was longer than 38 seconds (Mean = 17.1 s, SD = 9.2 s). There were also no limits on how many correspondences could be selected, except that only one correspondence was allowed for each mentation file.

RESULTS

A total of 20 correspondences were selected. Unfortunately, two of them came from a receiver who was excluded from the main experiment. This receiver had already participated in a Ganzfeld experiment, and one of the requirements for participation in the main experiment was no previous participation in a Ganzfeld experiment. However, as the hypothesis under investigation in the present study was not dependent on whether the receiver was a Ganzfeld novice or not, it was decided that those two correspondences should be retained. The selected correspondences are described in Table 1 below. The reader is urged to rate the degree of ‘impressiveness’ of each correspondence before continuing to the actual results.

Table 1. Descriptions of the 20 selected correspondences.

No	A frame from the clip	The mentation that was selected and the corresponding film clip
1		**Mentation**: *"A sun, a forest and a meadow. Butterflies. A child, or a woman who is chasing the butterflies. She is wearing a long, light-colored dress. From the turn of the [last] century." (In Swedish: "En sol, en skog och en äng. Fjärilar. Ett barn, eller en kvinna som jagar de här fjärilarna. Hon har lång ljus klänning på sig. Sekelskiftet.")* **Film clip**: Elvira Madigan, suicide pact. **Comment**: In the film clip library there were three clips from the movie Elvira Madigan. At several times, this receiver appeared to describe the content in all of these clips. However, the descriptions were seldom synchronized, i.e. they were seldom in real-time with the clip that was a part of the current set
2		**Mentation:** *"Also a feeling of something that explodes, fireworks or..." (In Swedish: "Också en känsla av att något exploderar, fyrverkeri eller...")* **Film clip:** Photographing Fairies **Comment:** This film clip is about a man who is coughing up fairies. Two assistants are trying to capture the event on film and use up a lot of magnesium powder flashes. The receiver talks several times about someone who is coughing and about explosions and fireworks.
3		**Mentation:** *"Falling." (In Swedish: "Falla.")* **Film clip:** Nature film, a diving eagle **Comment:** The receiver talks about "blue sky", "parachute jumping", "floating" and "bird". Exactly at the same time as the clip shows the eagle diving, the receiver says: "Falling".
4		**Mentation:** *"This sea animal is coming towards a little girl who is standing at the shore and she is quite petrified this little girl (In Swedish: "Det här sjöodjuret kommer fram emot en liten flicka som står vid sjökanten och hon är ganska så förstelnad den lilla flickan.")* **Film clip:** Night of the Hunter, hunted children **Comment:** This clip shows a man armed with a knife who is chasing two children, a girl and a boy. The children flee on a rowboat and the man is forced to stop pursuing them in the water, which makes him scream in anger.
5		**Mentation:** *"Shot." (In Swedish: "Skott.")* **Film clip:** Elvira Madigan, suicide pact **Comment:** This clip shows a woman who is chasing butterflies on a meadow. It ends with the picture being frozen followed by the sound of two shots being heard. And then the sound from two shots can be heard (it is the woman who is being shot). One second after the second shot, the receiver says: "shot".

6 

Mentation: *"Birds." (In Swedish: "Fåglar.")*
Film clip: The Birds, attacking birds
Comment: The receiver says many things that appear to describe the content of the clip. For example: At the same time as the film shows children being attacked by seagulls, the receiver says: "Birds".

7 

Mentation: *A chimpanzee (In Swedish: "En chimpans")*
Film clip: Nature film. Monkeys in a jungle.
Comment: The receiver says many things that fit this film clip. Besides "a chimpanzee" she says "the twitter of birds", exactly at the same time as the twitter of birds can be heard on the clip.

8 

Mentation: *"It is strange because I can feel it in my right forearm, in my right elbow. (In Swedish: "Det är märkligt för det känns i höger underarm, i höger armbåge.")*
Film clip: Photographing Fairies, a woman falls into a chasm.
Comment: The film clip is about a newly married couple who are taking a walk in the alps. Suddenly the woman falls into a chasm. Her husband succeeds in catching her hand and struggles to hold her. Eventually he loses his grip and she falls down into the chasm.

9 

Mentation: *"And there is a quarrel starting. I think that maybe weapons are being used." (In Swedish: "Och det bli nåt gräl som uppstår. Jag tror det är vapen inblandat.")*
Film clip: Reckless Kelly, a preaching cowboy.
Comment: The receiver says many things that fit well with this clip, e.g.: "Yes, it is clearly the wild west", "It reminds me of the Cartwright brothers[1]... in those cowboy hats. "

10 

Mentation: *"Where there's still blackness, there are flapping birds and there is mostly fog... and there are birds diving over the ocean, diving and catching something" (In Swedish: "Där det kvarvarande svarta är flaxande fåglar och så är det mest dimma... och det är fåglar som dyker över havet, dyker ner och fångar nånting")*
Film clip: The Birds, attacking birds
Comment: The receiver talked about flapping birds also on another occasion, but at that time, no birds were shown on the clip.

[1] From the TV series *Bonanza*.

11

Mentation: *"I can feel it in my fingers and in my arms that I am holding something. It is round. (In Swedish: "Jag kan känna det i fingrarna och i armarna att jag håller i nånting. Det är runt.")*
Film clip: Barbara, rowing after a ship
Comment: The clip is about a woman who is trying to row to a ship on which her lover is going away. The receiver says several things that appear to fit this clip quite well. For example, the receiver says: "I can feel movements, the chair is rocking."

12

Mentation: *"An old steam engine with its whistle blowing" (In Swedish: "Ett gammalt järnvägslok, som tjuter")*
Film clip: Fried Green Tomatoes, a boy gets stuck on the tracks.
Comment: The receiver says this at exactly the same moment as an old steam engine shows up with its whistle blowing and starts to move towards the boy who is stuck.

13

Mentation: *"Looks like something is being lifted up, a tong holding something, a loop." (In Swedish: "Ser ut som om det hissas upp nånting, en tång som håller i nånting, en ögla.")*
Film clip: Freaks, the shortest man on earth.
Comment: The receiver says this at exactly the same time as this clip shows how a dwarf with a special tong lifts some things in a kitchen.

14

Mentation: *"Cowboy... No, it is a man sitting on a horse." (In Swedish: "Cowboy... Nej, det är en man som sitter på en häst")*
Film clip: Dances With Wolves, a man helps a wounded native American woman.
Comment: The receiver says a lot of thing, besides the statement above, that fit well with this clip. For example: "This is America a hundred years ago.", "Are there Indians too?", "Now he rides away.", and "It's a horseman."

15

Mentation: *"Now I am not in space. Now a seagull comes, flies, big seagull, white, grey wings, yellow beak, flies down." (In Swedish: "Nu är jag inte i rymden, nu kommer en fiskmås, flyger, stor fiskmås, vit, gråa vingar, gul näbb, flyger ner")*
Film clip: The Birds, attacking birds
Comment: The receiver also says other things that fit well with this clip, e.g.: "And all the seagulls..."

16

Mentation: *"Big ear, protect your ears - she says." (In Swedish: "Stort öra, skydda dina öron säger hon.")*
Film clip: Custer, battlefield with a lot of corpses.
Comment: The clip shows how native American women puncture the ears of **a** dead soldier with some sharp object. Besides the statement above, the receiver did not say much that appeared to describe the content of this clip.

17

Mentation: *"Now I want to think about blood, blood. But I can't see any blood, I just think about it." (In Swedish: "Nu vill jag tänka på blod, blod. Men jag ser inget blod, jag bara tänker på det.")*
Film clip: Jeanne d´Arc, miracle on the battlefield.
Comment: The receiver gives the statement above at the same time as a woman on this clip says "I have seen enough blood." On another occasion the receiver says: "Yes, war, maybe war. War. I can see a German helmet." All soldiers in this clip wear helmets.

18

Mentation: *"Up in the air, the hands are being held out, upwards, as if someone tries to embrace the whole world." (In Swedish: "Uppåtgående, händerna går ut, uppåt, som om nån vill omfamna hela världen ungefär")*
Film clip: Color Purple, reunion
Comment: This clip shows how a mother and her daughter reunite after having been separated for several years. They run towards each other while holding their arms up. Finally they meet and embrace each other.

19

Mentation: *"I can see seagulls, I can hear seagulls around me. This kind of mattress on the water, blue. An inflatable mattress . Empty. No one is on it." (In Swedish: "Jag ser måsar, jag hör måsar runt mig. En sån här madrass på vattnet, blå. En uppblåsbar madrass. Tom. Ingen på den.")*
Film clip: Jaws, people fleeing from the water.
Comment: A blue inflatable mattress is actually being shown on this clip (and you can hear seagulls), but the receiver talks about the mattress 15 seconds after it appears in the clip.

20

Mentation: *"A feeling of a waterfall, that was a little closer, like looking diagonally upwards, as if there was a bridge or something above." (In Swedish: "En förnimmelse av vattenfall, som var lite närmare, som om blicken ändå var riktad lite snett uppåt som om det gick bro eller nånting ovanför.")*
Film clip: Romancing the Stone, escape over a ravine with a suspension bridge.
Comment: Everything in the statement above fits with this clip. There are also several other things the receiver says that fit with this clip, e.g.: "glances down, looks steep" exactly at the same time as the woman in the clip glances down and it looks very steep.

Of the 20 correspondences, 6 were correspondences with the target clip (= 30%) and 14 were correspondences with any of the decoy clips. This result is almost exactly at chance level ($p = .383$, one-tailed exact binomial test).

Correspondences no. 6, 7, 8, 13, 16 and 20 were "hits", i.e., the clip that the mentation appeared to describe was actually the target clip, the remaining (1, 2, 3, 4, 5, 9, 10, 11, 12, 14, 15, 17, 18 and 19) were "misses", i.e., the clip that the mentation appeared to describe was actually one of the decoy clips in the set.

DISCUSSION

The results apparently give no support to the psi-hypothesis. Remarkable correspondences between Ganzfeld mentation and target content, which have so often been reported from other Ganzfeld experiments, were also observed in the experiment under investigation in the present study. However, remarkable correspondences were also observed between Ganzfeld mentation and the content of the decoy clips and in fact, the distribution of the correspondences on target and decoy clips was very close to chance level.

So perhaps the remarkable correspondences between Ganzfeld mentation and target content observed in the present and previous studies are not so very remarkable after all. Perhaps they are altogether or mostly a question of subjective validation (a concept used by Marks, 2000, in explaining similar coincidental 'hits' both in experimental work and real life), Perhaps it is inevitable that experiments will show subjectively impressive correspondences between people's fantasies and movie contents if you try to look for such correspondences in material that contains about 30 hours of non-stop talking.

A quick look at the correspondences reported in Table 1 shows that in several cases the receiver has talked about water and about birds (especially seagulls). When judging all the mentations from the Ganzfeld experiment, the judge (JW) noted that water and birds were in fact very common themes, and in many cases, there was no water or birds on any of the four clips in the set to be judged.

It could be argued, however, that some of the remarkable correspondences obtained in Ganzfeld experiments were due to subjective validation, but that there is still a core of correspondences that cannot be explained away in this way. It could also be argued that the power of the current test was too low, as with 20 correspondences, as many as 9 (= 45%) would have to be 'hits' to reach significance. A possible way to get more precise and reliable measurements, and thus higher power, would be to let a group of people rate the 'impressiveness' of the 20 correspondences described in Table 1. This procedure was used in Study 2.

STUDY 2

In order to obtain greater statistical power, a group of students rated all 20 correspondences obtained in Study 1. The hypothesis (which was considered unlikely by the experimenter, JW) was that the mean rating of the 6 'hits' would be significantly higher than the mean rating of the 14 'misses'.

METHOD

Participants

Eleven undergraduate psychology students participated in the present study (7 women and 4 men, 27 - 41 years old) as part of course requirements (all the students at the psychology department at Stockholm university are required to participate a certain number of hours in studies conducted at the department, but they can freely choose which studies to participate in). All the students were taking a course in research methods and statistics, with the experimenter, JW, as the principal teacher. In this course, the students were required to read the discussion between Bem and Honorton (1994) and Hyman (1994), published in Psychological Bulletin. They were also required to participate in a seminar at which the tape by Kathy Dalton, described above, was shown and discussed. The advantage of having this population as judges was, of course, that the participants were all somewhat familiar with the Ganzfeld technique and with the concept of "remarkable correspondences between Ganzfeld mentation and film clip content". The disadvantage was that they all had met JW and thus could not be considered 100% blind to which mentation was a 'hit' and which was a 'miss' (even though the particular correspondences under investigation in the present study were not discussed, of course).

Material

Software and apparatus for presenting the 20 selected correspondences. The software used for the judging phase in Digital Ganzfeld was slightly adjusted so that it could present the 20 selected correspondences (see above) and was run on an IBM thinkpad r40 notebook computer, connected to a Liesegang dv 325 projector with a brightness of 1000 ANSI.

Rating form. The ratings of the 20 correspondences were made on a form on which the participants were to rate each correspondence on a scale from 0 – 100. The instructions on the form were that ”0” was to be used to denote a correspondence between the mentation and the content on the clip that was not impressive in the least, i.e. a correspondence that would very often turn up by chance alone. ”100” was said to denote a correspondence between the mentation and the content on the clip that was extremely impressive, i.e. a correspondence that would very seldom turn up by chance alone. The instructions also urged the participants to try to use the whole range of the scale, such that they rated at least one correspondence with a number close to 0 and at least one correspondence with a number close to 100, regardless of whether they thought that all correspondences were not at all impressive or whether they thought that all correspondences were extremely impressive. Those instructions were given in order to avoid a restriction of range problem in the statistical analysis.

Procedure

The experiment was conducted in a large classroom five minutes after a lecture that all participants had attended. They were told that they should read the instructions on the rating form, after which any questions would be answered. No questions were asked and the experimenter started the presentation software and left the classroom. A few seconds after the experimenter had left the room, the computer presented the 20 correspondences in a random order (using the Visual Basic 6.0 rnd-function, with the system timer as the seed). The participants had been instructed not to rate the correspondences the first time they were presented, as the purpose of the first presentation was to give the participants a feeling of the range of the degree of impressiveness of the correspondences. For each correspondence, the particular film clip was presented on a large screen, synchronized with the sound from the receiver in the Ganzfeld experiment. The exact wording of the receiver was also presented in written form under the film clip window, as it was sometimes difficult to hear the mentation. Between each presentation there was a 7-second long pause. After all the 20 correspondences had been presented once, the software presented a text for 10 seconds, stating that all correspondences were now to be presented once more, but that this time, they were to rate every correspondence. The software then presented the 20 correspondences in a new random order, but now with a 12-second pause between presentations. After all the correspondences had been shown again, the participants left the room and handed the forms to the experimenter. The experiment took 25 minutes to perform.

RESULTS

The 6 correspondences that were ’hits’ (correspondences between the mentation and the target clip in the Ganzfeld experiment) received a mean rating of 62.78 (SD = 7.82) and the 14 correspondences that were ’misses’ (correspondences between the mentation and one of the decoy clips in the Ganzfeld experiment) received a mean rating of 58.85 (SD = 18.94). An independent samples t-test with type of correspondence (hit, miss) as independent variable and mean rating of the correspondance as the dependent variable showed however that the difference was not significant: ${}_{.05}t_{18} = .485$, $p = .634$ (2-tailed). The mean ratings for all 20 correspondences, together with standard deviations and standard errors are presented in Table 2.

Table 2. Means, standard deviations and standard errors for the ratings of the 20 correspondences, sorted by means.

No	Hit/Miss	Mean	Std. Deviation	Std. Error
1	Miss	17.36	15.474	4.666
3	Miss	19.45	27.373	8.253
4	Miss	23.73	17.053	5.142
19	Miss	32.82	21.175	6.384
10	Miss	43.91	26.853	8.097
11	Miss	43.91	23.772	7.167
6	Hit	44.18	35.182	10.608
17	Miss	47.73	35.953	10.840
16	Hit	48.36	25.228	7.607
13	Hit	56.82	28.920	8.720
8	Hit	60.82	25.600	7.719
20	Hit	61.82	29.264	8.823
9	Miss	62.45	24.643	7.430
18	Miss	62.91	26.629	8.029
2	Miss	63.82	30.974	9.339
15	Miss	65.00	25.397	7.657
7	Hit	66.36	25.796	7.778
5	Miss	69.27	28.632	8.633
14	Miss	73.00	23.833	7.186
12	Miss	77.73	24.121	7.273

DISCUSSION

The correspondences that were 'hits' were not rated as significantly more impressive than were the correspondences that were 'misses'. The result is thus consistent with the notion that at least some of the remarkable correspondences so often reported to occur in Ganzfeld experiments are due to subjective validation. The result should be interpreted with some caution, however. The participants cannot be said to have been 100% blind, as they were all students of the experimenter.

GENERAL DISCUSSION

The present paper presents two studies, both of which have to be cross-validated to be conclusive. The first study showed that when a single judge blindly selected impressive correspondences between receiver mentation and film content, these correspondences were distributed between target and decoy clips almost exactly as would be expected by chance. The second study showed that a group of students did not rate the correspondences between mentation and target clips as significantly more impressive than the correspondences between mentation and decoy clips. The impression that the receiver sometimes can almost "see" the target, and because of this describe it in great detail thus appears to be due to a cognitive illusion after all.

The cognitive illusion theory is not the only possible conclusion, however. It could be argued that most or all of the 20 selected correspondences in Study 1 were due to psi. Following this explanation, the correspondences between mentations and decoy clips could be viewed as being due to clairvoyance, or, for those receivers who later watched all four clips in the used set, as being due to precognition. The same could be said for the correspondences between mentations and target clips, but perhaps now with telepathy as an additional explanation.

This kind of argument seems to be unfalsifiable, but that is actually not the case. It would be possible to take half of the 128 mentation files created in the Digital Ganzfeld study and scramble them so that they become associated with different sets than the original ones. A new judge could then be given the task of listening to all mentations, scrambled and unscrambled, and watching the clips in the associated sets and then trying to select the 20 most impressive correspondences. If significantly more of these correspondences belong to unscrambled mentations than to scrambled ones, the ”clairvoyance hypothesis” would gain support. Otherwise it loses credibility. As a matter of fact, a study following this procedure is presently being undertaken in Stockholm.

Finally, it should be noted that the 20 correspondences that were selected in study 1, were selected from data obtained in a digital ganzfeld study in which the main result measured as number of direct hits by an external judge was at chance level and by the receivers, significantly below chance level.

In comparing the quality of these correspondences with the ones described in the literature, it should also be emphasised that the methodology used here restricted the assessment to a single utterance per potential hit. Accordingly we could not test the hypothesis (Wright & Parker, 2003) that “best hits” can be identified by a multiplicity of specific real time correspondences, repeated themes, and unexpected changes in content corresponding to the film.

ACKNOWLEDGEMENT

We would like to acknowledge the financial support of John Björkhem Memorial Foundation.

REFERENCES

Bem, D.J., & Honorton, C. (1994). Does psi exist? Evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*, 4-18.

Goulding; A., Westerlund, J., Parker, A., & Wackermann, J. (2001). A new system for the automated digital Ganzfeld. *The Parapsychological Association 44th Annual Convention: Proceedings of Presented Papers*, 408-410.

Goulding; A., Westerlund, J., Parker, A., & Wackermann, J. (2004). The first real-time digital autoganzfeld study. Manuscript in preparation.

Honorton, C., Berger, R.E., Varvoglis, M.P., Quant, M., Derr, P., Schechter, E.I., & Ferrari, D.C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology, 54*, 99-139.

Hyman, R. (1994). Anomaly or artifact? Comments on Bem and Honorton. *Psychological Bulletin, 115*, 19-24.

Marks, D. (2000). *The psychology of the psychic.* Buffalo, New York: Prometheus.

Morris, R.L., Dalton, K., Delanoy, D.L., & Watt, C. (1995). Comparison of the sender / no sender condition in the ganzfeld. In *The Parapsychological Association 38th Annual Convention: Proceedings of Presented Papers, (pp. 244-259).* The Parapsychological Association.

Palmer, J. (2003). ESP in the Ganzfeld: Analysis of a debate. *Journal of Consciousness Studies, 10*, 51-68.

Parker, A. (2003). We ask, Does psi exists? But is it the right question and do we really want an answer anyway? *Journal of Consciousness Studies, 10*, 111-134.

Parker, A., Persson, A., & Haller, A. (2000). Using qualitative Ganzfeld research for theory development: Top down processes in psi-mediation. *Journal of the Society for Psychical Research, 64*, 65-81.

Wright, T, & Parker, A. (2003). An attempt to improve ESP scores using the real time digital Ganzfeld technique. *European Journal of Parapsychology, 18*, 69-76.

Address for correspondence: Joakim Westerlund,
Department of Psychology, Stockholm University,
S-106 91 Stockholm.
E-mail: joakim.westerlund@mbox200.swipnet.se

A TEST OF PREDICTIONS FROM FIVE STUDIES ON TELEPATHIC GROUP COMMUNICATION OF EMOTIONS

Joakim Westerlund, & Jan Dalkvist
Department of Psychology, Stockholm University

ABSTRACT

The purpose of the present study was to test a set of predictions from a previous series of telepathy studies, involving 337 university students. In these studies, groups of receivers were asked to discriminate between positive and negative emotional slide pictures that were being looked at by groups of senders. The senders and the receivers were sequestered in separate acoustically insulated rooms. None of the eight predictions was confirmed by data from a replication study, involving a total of 605 university students as participants. Discussed are three explanations denying that any telepathic communication occurred, either in the original series of experiments or in the new ones, and three explanations assuming that some telepathic communication did occur, after all. The three explanations denying the occurrence of any telepathic effect were: (1) All significant results obtained in the original experiments were caused by random variation. (2) The replication experiments were better controlled than the original experiments. (3) The original positive results were obtained through systematic selection. All of these three explanations were judged to be plausible. The three explanations assuming that some telepathic communication did occur were: (1) Some of the eight hypotheses were only partly true and need to be modified. (2) The distribution(s) of some critical person or situation variable(s) had changed during the six years of experimentation, which, in turn, has affected the results. (3) The replication experiments were run at less favourable times with respect to some physical factors: local sidereal time (LST) – an astronomical time and space measure, which is indirectly related to the magnitude of cosmic radiation that reaches the earth – and disturbances in the earth´s magnetic field, as measured by the ap-index. On the basis of empirical tests, the two latter explanations (changing distributions of critical personal or situational variables and changing LST or ap-index) were rejected as unlikely. It was concluded that, taken together, the three explanations assuming that no telepathic communication had occurred could well account for the failure to confirm any of the predictions.

INTRODUCTION

Since the spring of 1993, a series of telepathy studies has been performed at the Department of Psychology at Stockholm University, with one of us (JD) as initiator. The studies have all been concerned with transmission of emotions and have been carried out as group experiments.

The series of studies consists of two parts, an initial part, containing five studies (Studies 1-5), which mainly served to generate a set of hypotheses, and a second part, in which these hypotheses were tested in a replication study. The five original studies have been published elsewhere (Dalkvist & Westerlund, 1998a,b). A less detailed description of these studies will be given below.

The basic design of the present series of studies rests on two different ideas. One was to try to strengthen "weak" telepathic signals by using several senders instead of one. (Even if this idea should turn out to be wrong, using several participants has the practical advantage of permitting large amounts of data to be collected during a relatively short period of time.)

In spite of its practical advantages, group ESP experiments are rare, however. One reason may be Rhine's (1947/1971) negative evaluation of group testing as compared to individual testing:

> One of the discoveries made during these years was the great disadvantage of group testing compared to individual tests. Vernon Sharp and Dr. C.C. Clark of New York University compared classroom test with private tests and found, as we did at Duke, that individual tests give much better scoring rates. In fact, the group tests were little above the average expected from pure chance.... (p. 40)

Nevertheless, quite impressive results from group experiments on ESP have been reported, most notably by Carpenter (e.g., 1988).

Another reason why group experiments on ESP are relatively rare may be a statistical problem that is met in group testing: due to the possible occurrence of dependency among the participants´ responses in group testing, the statistical assumption of independent measures runs the risk of being violated (the so-called stacking problem). There are several ways of overcoming this problem, however, for example by simulation, a technique that has been used in the present series of studies as a complement to conventional statistical methods.

The other idea on which the present design is based is that strong emotional messages – signals of danger, for instance – may be, for evolutionary reasons, easier to transmit telepathically than are more neutral messages. Given the obvious link between emotions and needs, the notion that emotional stimuli are more efficient as ESP targets than are neutral ones is implied by evolutionary theories of ESP, such as those of Stanford (1978, 1982) and Taylor (1999), positing that ESP serves the function of helping the individual to fulfill his or her needs. A direct test of the notion that emotional stimuli facilitate ESP as compared to neutral stimuli was performed by Moss and his co-workers some decades ago (Moss & Gengerelli, 1969; Moss, 1969; Moss Chang & Levitt, 1970), using a design that was quite similar to the one used in the present series of studies (although individual senders and receivers, and not groups of senders and receivers, were used). The results supported the notion that emotionally arousing stimuli favour ESP performance. An attempt by Gelade and Harvie (1975) to replicate these findings did not succeed, however.

THE ORIGINAL STUDIES

A total of 337 participants, 222 females and 115 males, with a mean age of 26.7 yrs (range: 18-65 yrs) took part in the five original studies (Dalkvist & Westerlund, 1998a,b). Except for some 20 participants who came from other disciplines and were paid for their participation, participants were undergraduate psychology students at the Stockholm University, who chose to participate in the study as part of their course requirements. The studies comprised from 2 to 9 single experiments each, twenty-four in all; the mean number of participants per experiment was 13.75.

As stimuli, 30 slide pictures, 15 with positive motifs (e.g., nature pictures and pictures of happy people) and 15 with negative ones (e.g., pictures of traffic accidents or starving children) were used (for a complete description, see Dalkvist & Westerlund, 1998a,b).

When the participants arrived at the laboratory, they were (quasi)randomly divided into two groups, one sender group and one receiver group. The senders and the receivers were sequestered in two acoustically insularated rooms, with one room between them. The two experimental rooms were connected with each other by a signal device; a lamp in the receiver room could be turned on and off from the sender room. There were two experimenters in the sender room and two in the receiver room.

Before the experiment started, the participants rated their belief in telepathy on a three-point scale. In Experiments 4 and 5, corresponding ratings were also made after the experiment, but on a seven-point scale, to obtain more fine-grained data (we considered the possibility of substituting the three-point scale used before the experiment with a seven-point scale, but refrained from doing so as that would have made it more difficult to make comparisons).

The slides were presented in random orders, a new order for each group of senders. The senders´ only task was to look at the pictures and to “hold on to” the feelings evoked by the respective pictures as long as they were being shown. The receivers were instructed to guess whether a given picture was positive or negative (they were informed about the number of slides, but not that the number of positive and negative pictures was the same). One of the experimenters in the receiver room watched the signal lamp and reported to the receivers when a new picture was shown to the senders. Each picture was shown for 20 seconds, with an inter-stimulus interval of about 0.5 seconds.

When all 30 pictures had been shown, the participants changed rooms, and those who had served as senders now served as receivers and vice versa. Thus, each experiment consisted of two separate sessions, a first one in which half of the participants started as senders and the other half as receivers and a second session in which the roles were reversed, using the same stimulus pictures presented in a different random order.

With a view to obtaining a psychological description of the pictures used, they were rated on six different scales (by participants other than those taking part in the main experiments). All of the pictures were rated with respect to how (a) unpleasant/pleasant, (b) involving, (c) familiar and (d) perceptible (easy to apprehend) they were. In addition, the 15 negative pictures were rated with respect to how (a) compassion-arousing and (b) repulsive they were, and the 15 positive pictures with respect to how calm or exciting they were. Moreover, on the basis of the four "emotional" scales for negative pictures, a scale of negative emotionality was constructed.

Hit rate, defined as number of correct responses or proportion of correct responses (when stimulus data were analysed) was invariably used as the dependent variable in the data analyses. (No majority vote analyses have been done so far.) Hit rate was analysed as (a) a function of person or situational factors (for example, belief in telepathy and the order of the task as sender and the task as receiver) and (b) a function of stimulus factors (for example, rated characteristics of the pictures).

Data analyses were made both by means of conventional statistical methods and by means of a so-called Monte Carlo method, a simulation technique, alluded to in the introduction, which in contrast to conventional statistical methods does not require that any particular statistical assumptions be satisfied.

Together, the results from the five original studies seemed to support the hypothesis that telepathy exists. For example, several significant results were obtained when analysing effects of person and situational factors, the most consistent finding being that participants who believed in telepathy had a lower hit rate than that expected by chance. With respect to stimulus factors, hit rate for the first picture presented was higher than expected by chance. Furthermore, a negative correlation was obtained between negative emotionality and hit rate for the pictures with negative motifs; that is, the more unpleasant a negative picture was, the lower the hit rate tended to be. (No relationship was found between hit rate and emotional ratings for positive pictures.) Still another stimulus effect obtained was that the picture immediately preceding the current picture tended to affect the hit rate of that picture. For a more detailed presentation of the results from the original studies, see Dalkvist and Westerlund (1998b).

There were some problems with the results, however. The major problem was that, for the most part, the studies were explorative in character and had not been planned to test particular hypotheses. Thus, except for an initial hypothesis of an (unspecified) effect of belief in telepathy on hit rate, all the findings were made post hoc during the course of the studies.

THE REPLICATION STUDY – TEST OF PREDICTIONS

Procedure

On the basis of the results of the five above studies, a number of predictions were formulated and tested in a replication study, originally consisting of two parts. These predictions were all based on statistically significant (or, in one case, marginally significant) results obtained when data from the five studies were put together. The predictions (given below) were announced in *the Journal of Parapsychology* (Dalkvist & Westerlund, 1998b) before the data analyses had started.

The new study was an exact replication of the latest of the five original studies, except that two minor further control measures were adopted. The first one concerned the communication between one of the experimenters in the sender room and one of the experimenters in the receiver room. Previously, the experimenter in the sender room had orally informed the experimenter in the receiver room when the experiment was going to start. Because the experimenter in the sender room had just finished changing the order of pictures in the magazine and could then have noticed whether the first picture to be presented was negative or positive, it is conceivable (though highly unlikely) that this information had been unconsciously transmitted to the experimenter in the receiver room by means of body language or by the voice. This was now prevented by replacing the oral communication with communication by means of a light signal.

The second control measure prevented a conceivable, but also highly unlikely, possibility that some of the experimenters would see the random orders before the experiments. In the previous studies, the random orders were placed in an *open* envelope, in the room between the two experimental rooms (Studies 4-5) or in the experimenter room (Studies 1-3). In the replication study, the random orders were instead placed in a *sealed* envelope, between the two experimental rooms, and it was not opened until the pictures were going to be rearranged. The random orders were put into the envelope by JD, who was blind to their content, and the envelope was opened by JW.

The first part of the replication study was carried out in the autumn of 1997 and in the spring of 1998. Three hundred and thirty-five participants took part in this part of the study. Preliminary analyses suggested, however, that this number of participants might be too small for a critical test of the predictions to be made (we were uncertain about whether our failure to confirm any prediction might be due to sampling error). We therefore decided to increase our data set, using exactly the same procedure as in the first part of the study. This was done in 1999, with 270 new participants. (We were fully aware of the problem of optional stopping of data collection that was associated with this procedure.)

The total group of 605 participants comprised 432 females and 173 males, the mean age being 26.45 yrs (range: 18-75 yrs). The large majority of the participants were undergraduate psychology students at the Department of Psychology at Stockholm University, who chose to participate in the study as part of their course requirements; the remaining participants came from other disciplines and were paid for their participation.

There were 47 single experiments in all. The number of participants per experiment ranged from 5 to 19, with a mean number of 12.87.

As far as we can see, the data from the replication study are completely free from any conceivable systematic error.

Predictions and results

A total of eight predictions were tested. Five of them concerned person or situational variables. Four of these predictions concern effects obtained in a three-way ANOVA, with belief in telepathy as measured before the experiment, gender and sender/receiver order as independent variables, repetition aversion (defined as the number of times the subject shifted from one type of response, i.e. "positive picture" or "negative picture", to the other), number of positive responses and age as covariates, and hit rate as the dependent variable:

P1.1. *An effect of belief in telepathy as measured before the experiment will be obtained, with a positive mean hit rate for undecided participants, a negative mean hit rate for those participants who believe in telepathy, and a weak positive mean hit rate for participants who do not believe in telepathy.*

As can be seen from the table below, the direction of the prediction tended to be supported, although all mean hit rates were around the chance score of 15.

Table 1. Means, standard deviations and results from one sample t-tests (2-tailed) of deviation from 15 hits (expected by chance) for the three levels of belief in telepathy as measured before the experiment.

Belief in Telepathy	Mean	St. Dev.	t(df)	Sig.
"no"	15.10	2.84	.408 (125)	n.s.
"don't know"	14.98	2.77	-.154 (350)	n.s.
"yes"	14.91	2.27	-.450 (115)	n.s.

P1.2. *An interaction effect between gender and sender/receiver order will be obtained, with a positive mean hit rate for the males when they start as receivers, and a positive mean hit rate for the females when they start as senders.*

No significant interaction effect was obtained between gender and receiver order ($F_{1,597} < 1$): Males who started as receivers still performed better than males who started as senders, but this was also true for females. Thus, instead of a significant interaction effect between gender and receiver order, as predicted, a significant main effect of receiver order was obtained in an one-way ANOVA ($F_{1,\ 597} = 4.605$; $p < .05$), with a higher hit rate for those participants who started as receivers than for those who started as senders. When entered into the above-mentioned three-way ANOVA, however, receiver order was not significant ($F_{1,\ 578} = 2.354$; $p = .126$).

P1.3. *An effect of repetition aversion will be obtained, with a negative correlation between hit rate and repetition aversion.*

In the original studies, repetition avoidance was significantly related to hit rate when entered as a co-variate in the above-mentioned three-way ANOVA ($F_{1,\ 310} = 3.889$; $p < .05$), corresponding to a Pearson correlation of $r_{330} = -.128$ ($p < .05$). The F-ratio dropped to $F_{1,\ 578} = .026$ (n.s.) and the Pearson correlation to $r_{599} = -.006$ (n.s.) in the replication study.

P1.4. *An effect of age will be found, with a positive correlation between hit rate and age.*

When occurring as a covariate in the above-mentioned three-way ANOVA in the original studies, age exhibited a significant effect ($F_{1,\ 310}$; $p < .05$). The corresponding Pearson correlation was only marginally significant ($r_{330} = .097$; $p = .079$), however. The F-ratio dropped to $F_{1,\ 578} = .027$ (n.s.) and the corresponding Pearson correlation to $r_{599} = .011$ (n.s.) in the replication study.

The final prediction for person or situational variables is about belief in telepathy as measured after the experiment:

P2. *A negative correlation will be obtained between hit rate and belief in telepathy as measured after the experiment.*

In the original studies, a Pearson correlation of $r_{233} = -.185$ ($p < .01$) was obtained between hit rate and belief in telepathy as measured (on a 7-point scale) after the experiment. This correlation dropped to $r_{598} = -.034$ (n.s.) in the replication study.

One of the three predictions about effects of the stimulus condition is concerned with the position order of the pictures:

P3. *The first presented picture will exhibit a hit rate above that expected by chance.*

In the original studies, the first presented picture exhibited a hit rate of 58 % (Chi-square = 8.679; df =1; $p < .01$). In the replication study, hit rate dropped to 51.2 % (Chi-square = .372; df = 1; n.s.).

We have speculated that a more general decline effect (for instance, a linear decline effect or a decline effect formed as a ski slope) might be established when more data were available. We were not able to find any decline effect at all, however.

The next prediction was concerned with psychological characteristics of the pictures:

P4. *For negative pictures, a negative correlation will be obtained between hit rate and negative emotionality, as defined by combining the scales of pleasure, involvement, compassion, and repulsion.*

In Studies 2-5, a Pearson correlation of r_{13} = -.581 ($p < .05$) was obtained between hit rate and negative emotionality, defined as in the prediction (information for calculating negative emotionality was missing in Study1). In the replication study, the correlation dropped to r_{13} = -.088 (n.s.).

The final prediction was concerned with a particular sequence effect:

P5. *An effect of type of previous picture will be obtained, with a higher mean hit rate for a positive previous picture than for a negative one, irrespective of whether the present picture is positive or negative.*

The difference obtained in the original study, corresponding to a paired samples t-value of t_{332} = 2.192 ($p < .05$; two-tailed), was completely gone in the prediction study (t_{604} = -.015; n.s.; two-tailed). The prediction was thus disconfirmed.

In sum, none of the eight predictions was confirmed.

POSSIBLE EXPLANATIONS FOR THE NEGATIVE RESULTS

There are several different conceivable explanations for why the predictions were not borne out. Three of the six possible explanations considered below deny that any telepathic communication occurred, either in the original studies or in the replication study. The remaining three explanations presume that such effects did occur, after all.

The following three interpretations imply that no telepathic effects occurred:

1A. *We were unlucky: all significant results obtained in the original studies were caused by random variation.*

A necessary condition for considering this explanation as probable is that the different significant results were statistically dependent. On the contrary assumption, that the different tests were statistically independent, it is highly unlikely that all of the results are attributable to random fluctuations, because of the large number of significant results that were obtained. But if one assumes that significant effects from different analyses were strongly statistically dependent, the likelihood that the results are attributable to random variation increases. The probability would at most – when all dependencies are at maximum – roughly correspond to the significant p-values typically obtained in the original studies, that is, it would lie between 1% and 5%. Results that appear by chance that often may, of course, be questioned, particularly in an area such as parapsychology. A reasonable guess would be that the probability in question would fall somewhere between 1% and 1‰.

An argument favouring the interpretation now discussed is that the effects obtained in the original study were generally rather small.

2A. *The replication experiments were better controlled than the previous experiments.*

This explanation means, among other things, that the additional control measures adopted in the replication study – a protection against the possibility that some experimenter could have known the picture orders beforehand and elimination of all oral communication between the experimenters in the sender room and those in the receiver room – eliminated the cause or causes of the original positive results. This explanation also means that the positive results obtained in the three first, less well-controlled, studies (which corresponds to about 1/3 of the original data) could at least in part have been due to experimental errors, which were controlled for in subsequent studies.

On the present explanation, several apparently harmless experimental errors could together have resulted in the positive results of the original studies. This possibility is in line with the fact that the results tended to become increasingly weaker as the experimental control was sharpened. Alternatively, there was only one or a few, apparently innocent, but in reality fatal experimental errors.

In any case, if the present explanation is correct, this is highly remarkable and of great methodological interest – not only for parapsychology but for traditional psychology as well. The uncontrolled cues that the participants had been using must have been extremely subtle and perceived at an extraordinarily low unconscious level. Such possible cues are in general ignored in traditional experimental or other psychological research (but less often in parapsychological research). But if the explanation now considered is true, controls of various subtle cues should be adopted more generally.

3A. *The original positive results were found by systematic selection.*

It should first of all be pointed out that this explanation cannot hold for the effects obtained for belief in telepathy, as this variable was adopted already in the very first of the five original studies. Other positive results could well have been due to a systematic selection of positive results, however. For example, we would probably not have got the idea of analysing data for the first presented stimulus separately had the "first stimulus effect" not appeared as strongly as it did in our plots of hit rate against stimulus order. On the other hand, variables such as gender, age and sender/receiver order were *not* selected because they were found to give positive results, but because it seemed natural to look at them.

It seems unlikely that any of the above three explanations alone can account for the positive results in the original experiments. Taken together, however, they may well do so.

According to the following explanations, telepathic communication did occur, after all.

1B. *Some of the eight hypotheses were only partly true and need to be modified.*

An example of a possible necessary modification of such a hypothesis is the predicted interaction effect between gender and receiver order. This may have to be modified to a prediction about a main effect of receiver order according to the present results.

2B. *The distribution(s) of some critical person or situational variable(s) had changed during the six years of experimentation, which, in turn, has affected the results.*

One variable that we thought might have undergone a change was belief in telepathy. As a possibility, we thought that belief in the existence of telepathy might have increased among our students during the time of experimentation, perhaps as a consequence of serious parapsychological research becoming better known (due, maybe, to the attention that the present project has attracted).

However, our rating data do not indicate that any substantial change in belief in telepathy had taken place. As before, about 25% of the participants fell in the category "yes", about 50% in the category "don´t know" and about 25% in the category "no" on the 3-point "belief in telepathy" scale used before the experiment. The only noticeable tendency was a weak increment in the "don´t know" category and a corresponding decrease in the "yes" category, which might be taken to mean that our participants´ belief in telepathy had decreased somewhat rather than increased. It seems thus very unlikely that there was any change in belief in telepathy that would have changed the results in any significant way.

Like belief in telepathy, none of the remaining person or situational variables showed any change that could reasonably explain our failure to replicate the original results.

3B. *The replication experiments were run at less favourable times with respect to some physical factors.*

In a large number of different studies, Spottiswoode and May (1997) found a relationship between effect size in parapsychological experiments and so-called local sidereal time (LST) – an astronomical time and space measure, which is indirectly related to the magnitude of cosmic radiation that reaches the earth and, in turn, gives rise to disturbances in the earth's magnetic field. It is thus conceivable that our replication experiments were performed during less favourable physical circumstances than were the original experiments.

This possibility was tested. This was done for all studies, except for the first three, for which necessary dates were missing. Values of the *ap* geomagnetic index for all 3-hour intervals between 1994 and 1999 were collected from World Data Center C1 for Geomagnetism on the World Wide Web. For each single hour, the *ap*-index for the 3-hour interval encompassing the current hour was used. Because the distribution was markedly skewed, the values were logarithmized. The times for the *ap*-values were displaced one hour forward in time, in order to correct for the difference in time between UTC and Swedish time.

There was, in fact, a highly significant difference ($t_{843} = 23.92$; $p < 10^{-11}$) between the previous experiments and the replication experiments with respect to the average *ap*-index during the times of data collection, with lower values of *ap*-index during the replication experiments than during the original ones. However, there was no correlation between *ap*-index and hit rate. Likewise, there was no significant interaction effect between *ap*-index and any person or situational variable, as indicated by two-way ANOVA:s with hit rate as the dependent variable. These findings suggest that variations in *ap*-index cannot account for our failure to replicate the original positive results.

Based on the findings of Spottiswoode and May (1997), showing particularly good performances around 13:30 h LST, LST was divided into two intervals, one interval of "good" LST, ranging from 11:00 h LST through 16:00 h LST and one interval of "bad" LST, covering all other hours.

There was no significant difference between the time periods of the original studies and those of the new ones with respect to mean LST on the "good"/"bad" LST scale.

CONCLUSION

The major results from the present study are unusually clear: we were not able to confirm even one of the eight predictions being tested. In other words, the results from the original experiments could not be replicated, not even partially.

The most plausible general interpretation of this failure is that no telepathic communication had occurred in the present series of studies, neither in the original studies nor in the replication study, and that the discrepancy between the original studies and the new one must be attributed to statistical fluctuations or bias in the original studies. Perhaps Rhine (1947/1971) was correct in his evaluation that group testing is inferior to individual testing in ESP research (or perhaps telepathy does not exist).

With regard to the no-ESP interpretation, there is probably more than one explanation for the presumed false positive results in the original studies. A major probable explanation is selection of positive results, commonly called "data-snooping". Another probable explanation is that the three first studies were insufficiently controlled. In addition, chance may have played a joke on us. Together, two or all of these three explanations may well account for our failure to replicate the original findings.

The outcome of the present series of experiments may be seen as a typical example of a well-known and much debated phenomenon within parapsychology: the so-called decline effect, meaning in this context that initial studies of a certain type give positive results, which thereafter become weaker or disappear completely as attempts are made to replicate the initial positive findings (e.g., Milton & Wiseman, 1999) Among parapsychologists, a common attempt to explain this decline effect is to assume that researchers tend to lose their initial motivation as the same study is repeated, and, as a consequence, also their ability to create a psi-conducive mood on the part of the participants. (For sceptics, on the other hand, the decline effect proves that psi does not exist.) We do not believe that the declining motivation explanation is applicable to the present series of experiments, partly because we changed some of the experimenters during the original experiments as well as during the new ones.

ACKNOWLEDGEMENT

We would like to acknowledge the financial support of John Björkhem Memorial Foundation.

REFERENCES

Carpenter, J.C. (1988). Quasi-therapeutic group processes and ESP. *Journal of Parapsychology, 52*, 279-304.

Dalkvist, J., & Westerlund, J. (1998a). Four experiments on telepathic group communication of emotions. *Journal of Scientific Exploration, 12,* 583-603.

Dalkvist, J., & Westerlund, J. (1998b). Five experiments on telepathic communication of emotions. *The Journalof Parapsychology, 62*, 219-253.

Gerlade, G., & Harvie, R. (1975). Confidence ratings in an ESP task using affective stimuli. *Journal of the Society for Psychical Research, 48*, 209-219.

Milton, J., and Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125,* 387-91.

Moss, T., & Gerandelli, J.A. (1968). ESP effects generated by affective states. *Journal of Parapsychology, 32*, 90-100.

Moss, T. (1969). ESP effects in "artists" contrasted with "non-artists," *Journal of Parapsychology, 33*, 57-69.

Moss, T. , Change, A.F., & Levitt, M. (1970). Long-distance ESP: a controlled study. *Journal of Abnormal Psychology, 76*, 288-294.

Rhine, J. B. (1947/1971) The Reach of the Mind. New York: William Sloane.

Spottiswoode, J., & May, E. (1997). Anomalous cognition effect size: Dependence on sideral time and solar wind parameters. *Proceedings of the 40th Annual Convention of the Parapsychological Convention,* 399-409.

Stevens, P. (2003). A feedback-reinforcement model. *Proceedings of the Parapsychological Association 45th Annual Convention,* 273-81.

Stanford, R.G. (1978).Toward Reinterpreting Psi Events. *Journal of the American Society for Psychical Research, 72*, 197-214.

Stanford, R.G. (1982). An experimentally testable model for spontaneous extrasensory events. In I Grattan-Guiness (Ed.), Psychical research: *a guide to its history, principles and practices* (pp. 195-205). Wellingborough, UK: Aquarian Press.

Taylor, R. (1999). Evolutionary Theory and Psi: Reviewing and Revising some Need Serving Models in Psychic Functioning. *Proceedings of the Parapsychological Association 42nd Annual Convention,* 398-414.

Address for correspondence: Joakim Westerlund,
Department of Psychology, Stockholm University,
S-106 91 Stockholm.
E-mail: joakim.westerlund@mbox200.swipnet.se

'IT'S STILL BENDING': VERBAL SUGGESTION AND ALLEGED PSYCHOKINETIC ABILITY

Richard Wiseman, & Emma Greening
Psychology Department, University of Hertfordshire

ABSTRACT

Some alleged psychics appear to be able to deform metallic objects, such as keys and cutlery, by thought alone. This paper describes two studies that examined whether one aspect of these demonstrations, often referred to as 'after-effects', could be created by verbal suggestion. An after-effect occurs when a metal object is placed on a table but apparently still continues to bend by a small but noticeable amount. Some individuals believe that such effects cannot be the result of trickery because the alleged psychic is no longer in contact with the object, and therefore view eyewitness reports of this phenomenon as strong evidence for the paranormal. In the first study, participants (N=46) were shown a videotape in which a fake psychic placed a bent key on a table. Participants in one condition (Suggestion) heard the fake psychic suggest that the key was continuing to bend, whilst those in the other condition (No Suggestion) did not. All participants were then asked to complete a questionnaire regarding the degree to which they believed the key continued to bend once it had been placed on the table. Participants in the suggestion condition were significantly more likely to report that the key continued to bend. A second study (N=100) replicated these findings. In addition, participants were asked to provide a short verbal description of the events depicted on the videotape. These descriptions were rated by two judges, and the results revealed that participants who believed that the key had continued to bend were significantly less likely to recall that the fake psychic had suggested the continued bending of the key. Both studies also examined the levels of confidence expressed by participants in the accuracy of their recall. In the first study, participants in both the Suggestion and No suggestion conditions displayed a high level of confidence in their recall of this aspect of the video. In the second study, participants who reported that the key continued to bend displayed a significantly higher level of confidence in their testimony than others. Both studies also examined the possible relationship between belief in the paranormal and suggestion. In both studies participants were asked to complete a questionnaire measuring the degree to which they believed in ESP and PK phenomena. Neither experiment revealed any differences between participants who expressed a prior belief in the paranormal compared to those that did not. The paper discusses the implications of these results for the psychology of suggestion, testing individuals claiming macro-PK abilities and the assessment of eyewitness testimony for anomalous events.

INTRODUCTION

Psychics and mediums claim to possess a diverse range of paranormal powers, including, for example, the ability to predict the future, communicate with the dead and read minds (Wiseman & Morris, 1995a). One of the most controversial claims is that of psychokinetic metal bending (PKMB) - the alleged ability to deform metallic objects, such as keys and cutlery, by thought alone. In a typical PKMB demonstration, a metal object appears to bend whilst being gently held by an alleged psychic. Psychologists, parapsychologists and magicians have all explored ways in which such demonstrations can be duplicated by trickery. Each group has approached the issue from a slightly different perspective, with psychologists focusing on the potential of such work to inform the psychology of deception (Marks and Kammann, 1980), parapsychologists concentrating more on its importance for assessing allegedly genuine demonstrations of PKMB (Randi, 1983a,b: Truzzi, 1987), and magicians examining how such work can help enhance their performances (Fuller, 1975: Harris, 1985).

The resulting literature has explored a wide variety of possible methods for faking PKMB, including, for example, the switching of straight objects for pre-bent duplicates, the concealed application of force and ways of secretly inducing metallic fractures (see, e.g., Fuller, 1975; Randi, 1975; Marks & Kammann, 1980; Harris, 1985). However, these types of methods do not account for an intriguing aspect of alleged PKMB that has come to be referred to as the 'after-effect'. In a typical PKMB demonstration, the alleged psychic apparently uses his or her paranormal abilities to deform an object. In some demonstrations the alleged psychic then produces an 'after-effect', wherein the object is placed on a table but apparently still continues to bend by a small but noticeable amount (see, e.g., Marks & Kammann, 1980: Hasted, 1981). Some individuals believe that such effects cannot be the result of trickery because the alleged psychic is no longer in contact with the object, and therefore view eyewitness reports of this phenomenon as strong evidence for the paranormal (see, e.g., Panati). In contrast, skeptics and magicians have argued that such effects could be created by verbal suggestion. For example, in a book devoted to methods for faking alleged PKMB, magician Ben Harris (1985) noted:

> If you are doing a really convincing job, then you should be able to put a bent key on the table and comment 'Look, it is still bending," and have your spectators really believe that it is. This may sound the height of boldness; however, the effect is astounding – and combined with suggestion, it does work. (p.46)

As a result of such speculations, some psychologists have argued that testimony describing apparent PKMB after-effects is not evidential of genuine psychic ability (see, e.g., Marks, 2000).

The idea that PKMB after-effects can be created by verbal suggestion hasn't been subjected to any form of systematic evaluation or investigation. This is unfortunate for two main reasons. First, from a psychological perspective, such research has the potential to help inform our understanding of suggestion. The majority of laboratory research into verbal suggestion involves participants being shown information about a relatively commonplace event (e.g., slides or a videotape of a minor car accident), receiving incorrect information about what they have just witnessed and then being tested on their recollection of stimuli material (see, e.g., Loftus, 1992; Roediger, Wheeler, & Rajaram, 1993; Loftus, 1997). However, very little research has examined the effects of verbal suggestion on the perception of an ongoing event, or events that are as unusual as alleged PKMB. As such, the results of such work could play an important role in expanding our understanding of the potential relationship between verbal suggestion, observation and recall. Second, from a parapsychological perspective, such work could help evaluate the reliability of testimony for alleged PKMB after-effects, and therefore the degree to which such testimony constitutes evidence for the paranormal. This paper addresses these issues by describing two studies that explored whether it is possible to create testimony for PKMB after-effects via verbal suggestion alone.

In the first experiment, participants were shown a videotape of an apparent PKMB demonstration. During the videotape a performer stroked the stem of a key and revealed that this had apparently caused the stem to develop a very noticeable bend. He then placed the key on the table. Participants in the Suggestion condition then heard the performer suggest that the key was continuing to bend. In contrast, participants in the No Suggestion condition saw exactly the same footage but did not hear this suggestion. Participants were then asked whether the key had continued to bend whilst it lay on the table. It was predicted that those who heard the suggestion would be significantly more likely to report that the key had continued to bend.

The study also investigated the role that participants' belief in the paranormal may play in mediating such effects. Psychologists have carried out a considerable amount of work into the psychology of paranormal belief (see, e.g., reviews by Irwin, 1993, 1999; French, 1992), with some researchers reporting a positive relationship between such beliefs and several different measures of suggestibility (Haraldsson, 1985; Wiseman, Greening and Smith, 2003; Hergovich, 2003). The experiments reported here extended this work by examining the possible relationship between paranormal belief and verbal suggestibility within the context of PKMB after-effects. Prior to seeing the videotape, participants completed a questionnaire about their belief in the paranormal, and it was predicted that participants who tended to believe in the paranormal would be more susceptible to the effects of suggestion than disbelievers.

EXPERIMENT ONE – METHODS

Participants

46 psychology undergraduates (19 male, 27 female: mean age = 21.34, s.d. = 4.33) participated in the study in return for course credit.

Videotapes

The videotape shown in the Suggestion condition lasted approximately two minutes. It began with a performer and an interviewer sitting at a table with several objects (cutlery, packs of cards, keys) in front of them. The interviewer briefly described the objects and asked the performer to select one of them. The camera shot then changed to a close-up of the performer's hands, and showed him picking up a key and apparently using his psychokinetic ability to place a 25 degree bend in its stem (in reality this bend was achieved by sleight of hand). The performer then placed the key back on the table, and the videotape ended with a sixty-second close-up shot of the bent key. This shot was completely stationary and the key did not continue to bend. The soundtrack to this footage carried a single comment from the performer: in which he suggested that the key was continuing to bend. The videotape used in the No Suggestion condition was identical to that used in the Suggestion condition, except that the performer's comment was removed from the soundtrack.

Questionnaires

Belief in the Paranormal Questionnaire (BPQ)

The BPQ contains six questions concerning whether some people possess different forms of extrasensory perception and psychokinetic abilities (e.g., 'Do you think that some people can, just by mental effort, apply a noticeable force to an object?'). Participants indicate their responses to each question on a seven-point scale from 1 (Definitely No) to 7 (Definitely Yes). Higher scores indicate a greater belief in the paranormal (see Wiseman & Morris, 1995b, for further details).

Fixed Response Questionnaire (Fixed-RQ)

The first three items on the Fixed-RQ consisted of statements about events on the videotapes. Two of the statements were filler items and referred to the interviewer touching the objects on the table, and the performer saying that the key was heating up as it apparently bent in his hands. The third item, which was used as the dependent variable in the study, stated 'After the key was placed on the table, it continued to bend'. The fourth item on the questionnaire asked participants whether they thought that the bending of the key was paranormal. For all four items, participants were asked to make their responses on a seven-point scale from 1 (Definitely No) to 7 (Definitely Yes). The midpoint of the scale was labeled 'Uncertain'. Participants were also asked to rate their level of confidence in each of their answers on seven-point scales from 1 (Not at all confident) to 7 (Very confident).

Procedure

Participants were run in small groups of up to five individuals. Each group was assigned to either the Suggestion or No Suggestion condition. Participants were informed that the study would involve observing a videotape containing some anomalous phenomena, but not told that the experiment was examining the effects of suggestion on observation and recall. All participants first completed the BPQ. Participants in the Suggestion condition were then shown the videotape in which the performer suggested that the key was continuing to bend after it had been placed on the table. Participants in the No Suggestion condition were shown the videotape in which this suggestion had been removed. Finally, all participants completed the Fixed-RQ.

RESULTS

Participants' scores on the BPQ were summed over the six items (mean=22.11, s.d.=6.72). Participants were then assigned to one of two groups (Believers or Disbelievers) on the basis of a median split. The Disbelievers (N=22) had a mean BPQ score of 16.64 (s.d.=3.86), whilst the Believers (N=24) had a mean BPQ score of 27.12 (s.d.=4.46).

A 2 x 2 factorial ANOVA examined the relationship between suggestion (Suggestion vs No Suggestion), belief (Disbelievers vs Believers) and participants' scores on Item 3 of the Fixed-RQ. This revealed a highly significant effect of suggestion (F (1, 42) = 16.90, p = .0002), a non-significant effect of belief (F (1, 42) = 0.02, p = .88) and a non-significant interaction (F (1,42) = 3.09, p = .09). Participants' scores in the Suggestion condition were significantly higher than participants' scores in the No Suggestion condition, indicating that they were more convinced that the key continued to bend whilst it lay on the table (see Table 1).

Table 1: Means and standard deviations (in parentheses) for responses to Item 3 on the Fixed-RQ by suggestion and belief.

	Suggestion	No Suggestion	Total
Believers	3.42 (1.93) N = 12	2.25 (1.14) N = 12	2.83 (1.66) N = 24
Disbelievers	4.36 (2.33) N = 11	1.45 (.93) N = 11	2.91 (2.29) N = 22
Total	3.87 (2.14) N=23	1.87 (1.10) N=23	

Post-hoc unpaired t-tests compared scores obtained in the Suggestion condition with those obtained in the No Suggestion condition for Believers and Disbelievers. The scores for Believers were not significant (df=22, t[unpaired]=1.80, p[2 tailed]=0.08), while the scores for Disbelievers were highly significant (df=20, t[unpaired]=3.84, p[2 tailed]=0.001). These analyses suggest that, contrary to expectations, Disbelievers were actually more suggestible than Believers.

To discover the percentage of participants who believed that the key continued to bend, participants were split into two groups on the basis of their responses to Item 3 on the Fixed-RQ. Those that responded with either a '5', '6' or '7' were allocated to the 'Key continued to bend' group, whilst those that responded with either a '1', '2', '3' or '4' were allocated to the 'Key did not continue to bend' group. As shown in Table 2, almost 40% of participants in the Suggestion condition reported that the key continued to bend, versus just 5% in the No Suggestion condition. A Chi-squared analysis between group and suggestion was highly significant (df = 1, Chi-Square [with continuity correction] = 6.26, p [2 tailed] = .01).

An unpaired t-test revealed no significant difference between the confidence ratings given to item 3 on the Fixed-RQ by participants in the 'Key continued to bend' group versus those in the 'Key did not continue to bend' group (df = 44, t-value [unpaired] = .18, p-value [2 tailed] = .86: See Table 2). Both groups expressed relatively high confidence in the accuracy of their answers (approximately 5.7 on a 7 point scale).

Table 2: Numbers, percentages and confidence data for participants in the 'Key continued to bend' and 'Key did not continue to bend' groups.

	Suggestion	No Suggestion	Confidence ratings given to item 3 on the Fixed-RQ means (and standard deviations)
Key continued to bend group	9 39.13%	1 4.35%	5.66 (1.07)
Key *did not* continue to bend group	14 60.87%	22 95.65%	5.69 (1.53)

DISCUSSION

This study examined whether it was possible to create PKMB after-effects via verbal suggestion. The results revealed that participants in the Suggestion condition were significantly more likely than those in the No Suggestion condition to report that the key was continuing to bend. The size of the effect was far from trivial, with approximately 40% of participants in the Suggestion condition reporting continued bending of the key compared to just 5% of participants in the No Suggestion condition. The results also showed that participants who reported that the key continued to bend were relatively confident that their testimony was reliable. In addition, the study explored the possible relationship between belief in the paranormal and suggestibility. Contrary to expectations, the findings demonstrated that participants who expressed a prior belief in the paranormal were no more or less likely than disbelievers to report that the key had continued to bend. Taken together, these findings support the notion that a relatively small amount of verbal suggestion can create the types of after-effects often reported by people who have witnessed demonstrations of ostensible PKMB. However, the study did not support the hypothesis that participants who believed in the paranormal are more susceptible to observing and reporting such effects than those who disbelieved in the paranormal.

Experiment Two built upon the first study in several ways. First, it aimed to replicate the findings obtained in Experiment One with a greater number of participants. Second, the study extended this work by examining whether participants who reported that the key was continuing to bend also tended to recall the performer suggesting that this was the case. The impetus for this aspect of the study came from a colleague who, after attending a presentation about Experiment One, noted that although some witnesses of PKMB after-effects did mention the alleged psychic stating that the object was continuing to bend, many did not. This could be explained in one of two ways. First, it is possible that the latter set of witnesses were accurate and that the alleged psychic did not suggest that the object was continuing to bend. If this were the case it seems far less likely that the PKMB after-effects described by these witnesses were due to verbal suggestion. Alternatively, it is possible that the alleged psychics did suggest the continued bending of the object, but that the witnesses forgot about these comments or did not think that it was important to report them. This latter interpretation would fit with previous research into the observation and reporting of fake psychic demonstrations. For example, Wiseman & Morris (1995b) asked observers to watch a videotape of fake psychic demonstrations (e.g., spoon bending and the guessing of ESP cards) and then recall the events on the videotape. Many observers failed to recall events that were central to the way in which the tricks were achieved (e.g., the fake psychic placing the spoon briefly out of sight), thus making it very difficult to accurately reconstruct and explain the tricks on the basis of their testimony. For this reason, in Experiment Two it was predicted that participants who reported that the key was continuing to bend would be less likely than others to report that the performer suggested this was the case. To test this hypothesis, it was decided not to ask participants a direct question about whether the performer suggested the key continued to bend, as this question could act as a cue that would help them accurately recall such comments, but rather use a more naturalistic approach and ask them to simply provide a qualitative description of the events on the stimuli videotape. As such, Experiment Two employed both the Fixed-RQ from Experiment One and a supplementary qualitative questionnaire.

EXPERIMENT TWO – METHODS

Participants

100 psychology and cognitive science undergraduates participated in this study (36 male, 64 female: mean age = 22.15, s.d. = 5.44) in return for course credit.

Materials

The videotapes, PBQ and Fixed-RQ were identical to those used in Experiment One. However, this experiment also used an additional qualitative questionnaire. This Free-Response Questionnaire (Free-RQ) consisted of two sections. The first section asked participants to provide a written description of what happened on the videotape from the time that the performer picked up the key and placed it down onto the table (i.e., the first bend). The second question asked: ‘Please describe everything that happened from the time the performer placed the key back down on the table to the end of the film’. Participants were asked to provide as much detail as possible when answering both questions.

Procedure

Participants were run in small groups of up to ten individuals. Each group was randomly assigned to either the Suggestion or No Suggestion condition. All participants first completed the BPQ. After seeing the relevant videotape, participants completed both the Fixed-RQ and Free-RQ. The order in which the questionnaires were administered was counter-balanced between participants.

RESULTS

As in Experiment One, participants’ scores on the BPQ were summed over the six items (mean = 21.77, s.d.= 8.58). Participants were then classified as either Believers or Disbelievers on the basis of a median split. The Disbelievers (N=49) had a mean BPQ score of 14.61 (s.d. = 5.12), whilst the Believers (N=51) had a mean BPQ score of 28.65 (s.d. = 4.73).

A 2 x 2 factorial ANOVA examined the relationship between Suggestion (Present and Absent), Belief (Disbelievers and Believers) and participants’ scores on Item 3 of the Fixed-RQ. As in Experiment One, this revealed a highly significant effect of suggestion ($F(1, 96) = 26.19$, $p = .0001$), a non-significant effect of belief ($F(1, 96) = 0.16$, $p = .68$) and a non-significant interaction ($F(1,96) = 0.14$, $p = .70$). Participants’ scores in the Suggestion condition were higher than participants’ scores in the No Suggestion condition, thus replicating the findings of Experiment One.

Table 3: Means and standard deviations (in parentheses) for responses to Item 3 on the Fixed-RQ by condition and belief.

	Suggestion	No Suggestion	Total
Believers	3.9 (2.26) N = 30	1.86 (1.06) N = 21	3.06 (2.11) N = 51
Disbelievers	3.91 (2.35) N = 22	2.15 (1.20) N = 27	2.94 (1.99) N = 49
Total	3.90 (2.28) N=52	2.02 (1.40) N=48	

Again, post-hoc unpaired t-tests compared scores obtained in the Suggestion condition with those obtained in the No Suggestion condition for Believers and Disbelievers. This time, both the scores for Believers and Disbelievers were significant (Believers: df=49, t[unpaired]=3.84, p[2 tailed]=0.0004. Disbelievers: df=47, t[unpaired]=3.39, p[2 tailed]=0.001).

Scores on item 3 of the Fixed-RQ were again used to allocate participants into the 'Key continued to bend' group and the 'Key did not continue to bend' group. As shown in Table 4, almost 36% in the Suggestion condition believed that the key continued to bend versus no participants in the No Suggestion condition. A Chi-squared analysis between group and suggestion was highly significant (df = 1, Chi-Square [with continuity correction] = 19.34, p [2 tailed] = .0001).

An unpaired t-test revealed that participants in the 'Key continued to bend' group were significantly more confident in their answer to item 3 on the Fixed-RQ than participants in the 'Key did not continue to bend' group. (df = 98, t-value [unpaired] = -2.01, p-value [2 tailed] = .05: See Table 4). The mean of both groups was again relatively high (approximately 5.6 on a 7 point scale).

Table 4: Numbers, percentages and confidence data for participants in the 'Key continued to bend' and 'Key did not continue to bend' groups.

	Suggestion	No Suggestion	Confidence ratings given to item 3 on the Fixed-RQ means (and standard deviations)
Key continued to bend	19 36.54%	0 0%	6.10 (1.24)
Key did not continue to bend	33 63.46%	48 100%	5.15 (1.97)

Free-RQ

Participants' responses to question two on the Free-RQ were coded along two dichotomous dimensions. The first dimension concerned whether participants reported that the performer suggested that the key continued to bend (coded as either 'yes' or 'no'). The second dimension focused on whether the participant reported that the key continued to bend as it lay on the table (coded as either 'yes' or 'no'). Each of the participants' responses were coded by two raters (inter-rater reliability = .90), with any disagreements settled by discussion. The resulting codings allowed participants' responses to be classified into one of four groups. These groups, along with typical answers from each group were as follows:

Reported performer's suggestion and that the key continued to bend.

'When the key was put on the table the performer...claimed it was still bending. This was difficult to see but it did appear to bend a little'.

Reported performer's suggestion but not that the key continued to bend.

'The performer commented that the key continued to bend, however, I failed to notice any difference – the key remained the same'.

Didn't report performer's suggestion but reported that the key continued to bend.

'The key, which had already bent, continued bending without being touched by the performer'.

Didn't report performer's suggestion or that the key continued to bend.

'The key was placed down on the table and the performer placed his hands on the table'.

In the No Suggestion condition, no participants reported the performer's suggestion (perhaps not surprisingly, given that this had been removed from the soundtrack of the videotape) and none of them reported that the key continued to bend once it had been placed on the table. As such, all responses were placed into the category: 'Didn't report performer's suggestion or that the key continued to bend'.

The number and percentage of participants falling into each of the four categories in the Suggestion condition is shown in Table 5 below.

Table 5: Number and percentage of participants (Suggestion only condition) in each of the four categories regarding whether they reported the performer's suggestions and the continued bending of the key.

	Reported that the key continued to bend	Didn't report that the key continued to bend	Total
Reported performer's suggestion	2 12.5%	27 75%	29 55.77%
Didn't report performer's suggestion	14 87.5%	9 25%	23 44.23%

The majority (almost 90%) of those participants who reported that the key continued to bend, did not report that the performer had suggested that this was the case. This pattern was reversed among those that did not report that the key continued to bend, with approximately 75% of these participants correctly reporting the performer's suggestion. A Chi-Square analysis revealed that this relationship was highly significant (df = 1, Chi-Square [with continuity correction] = 15.09, p [2 tailed] = .0001).

Post hoc Chi-Square analyses revealed that Believers were no more or less likely than Disbelievers to report that the key continued to bend (df = 1, Chi-Square [with continuity correction] = .51, p[2 tailed] = .47) or that the performer suggested this was the case (df = 1, Chi-Square [with continuity correction] = 1.55, p[2 tailed] = .21).

Discussion

This paper describes two studies that explored whether ostensible PKMB after-effects can be created via verbal suggestion alone. In both studies, participants in the Suggestion condition watched a videotape in which a performer stated that a bent key was continuing to bend, whilst those in the No Suggestion condition watched an identical videotape but did not hear the performer's comments. Despite the videotape clearly showing that the key remained completely stationary, participants in the Suggestion condition reported significantly more movement of the key than those in the No Suggestion condition. Both studies revealed that this effect was far from trivial, with approximately 40% of participants in the Suggestion condition reporting that the key had continued to bend, versus just a few percent of those in the No Suggestion condition. In the second study, participants who believed that the key had continued to bend expressed significantly greater levels of confidence in the reliability of their testimony than those who correctly reported that the key remained stationary. As such, the studies provide strong support for the notion that even a relatively small amount of verbal suggestion can cause a large number of people to confidently report a PKMB after-effect.

As noted in the Introduction, the majority of work examining the relationship between suggestion and observation has tended to examine the influence of post-event suggestion on the recall of relatively commonplace events. The two studies described here provide compelling evidence that verbal suggestion can also significantly influence the ongoing observation and subsequent reporting of an unusual, and apparently anomalous, event. It is hoped that these findings will encourage other researchers to employ this type of paradigm to further explore some of the many issues raised by this work. For example, although the effects obtained in the current studies appear both robust and large, it seems quite possible that even stronger effects could be obtained under more naturalistic conditions. In the experiments described here, observers watched a videotape within the context of a laboratory experiment. Future work could explore the effect of observing a live, rather than taped, performance, and also explore how other forms of context (e.g., the performer demonstrating a magic trick in a bar or an apparently genuine paranormal effect at a psychic fair) impacts upon people's observation and recall. In addition, the verbal suggestion in the present studies consisted of just a single sentence and was not endorsed by the interviewer on the videotape. Future work could explore the possible effects of presenting different amounts and forms of comments, and the impact of having authority figures and other observers endorse these suggestions. Finally, both of the present studies examined the effects of ongoing verbal suggestion on a rather unusual type of event (i.e., a PKMB demonstration). Future work could explore the extent to which the same type of suggestion may influence participants' observation and reporting of the types of stimuli more usually used in research on post-event suggestion (e.g., car accidents and crimes).

On a more pragmatic level, the findings reported here demonstrate that testimony for PKMB after-effects can be created by verbal suggestion, and therefore the testimony from individuals who have observed allegedly genuine demonstrations of such effects should not be seen as strong evidence in support of the paranormal. The work therefore emphasizes the need for those wishing to properly investigate such phenomenon to carefully film or videotape such demonstrations, rather than rely on eyewitness reporting.

Both studies also examined the possible relationship between participants' prior belief in the paranormal and the degree to which they reported the key continuing to bend. On the basis of past research it was predicted that participants who believed in the paranormal would be more likely than disbelievers to report that the key continued to bend. This hypothesis was not supported in either study (indeed, in the first study post-hoc tests suggested that Disbelievers may have actually been more suggestible than Believers), and thus the obtained results are consistent with the small number of studies that have not reported a relationship between belief in the paranormal and suggestibility (see, e.g., Saucer, Cahoon, Delwin and Edmonds, 1992). This finding could be interpreted in several ways. First, it is possible that the previous work may be incorrect and that, in reality, the relationship between belief in the paranormal and suggestibility is nonexistent or at least much weaker than originally thought. This might be the case if, for example, the small number of published studies examining this effect represent a biased sub-set of studies that had been undertaken, with the majority of unpublished studies not supporting the predicted relationship. Alternatively, it may be that the relationship between belief in the paranormal and suggestion is more complex than first thought. Suggestibility can be measured in many different ways. For example, Haraldsson (1985) employed a standard suggestibility scale frequently used in hypnosis research whereas Wiseman, Greening and Smith (2003) asked participants to assess the degree to which luminous objects were apparently moving within the context of a live séance. Both of these measures are clearly different from one another, and also differ from the measure used in the current two studies (i.e., estimating the degree to which a bent key was continuing to bend on a videotape). It seems quite possible that these very different measures reflect quite different forms of suggestion, and that individuals who believe in the paranormal may only perform differently to disbelievers on certain measures. These competing interpretations could be resolved by future research.

Experiment Two also assessed the notion that participants who reported that the key had continued to bend would be significantly less likely than others to report that the performer had suggested this was the case. The results supported this hypothesis. This finding provides additional support for the notion that eyewitness reports of allegedly anomalous events may omit the information needed to properly assess some of the potential normal explanations that may account for such events (see Wiseman & Morris, 1995b). This effect could be interpreted in several ways. It is possible, for example, that participants who believed that the key continued to bend, thought that they had witnessed a genuine effect (i.e., that the movement of the key was not illusory) and therefore the performer's comments at the time were not attended to, or later reported, because they seem relatively unimportant. This interpretation would be consistent with previous work from schema theorists examining how the way in which people interpret an event dictates the perceived importance of certain information about that event, and therefore the likelihood of them attending to, and recalling, such information (see, e.g., Anderson & Pichert, 1978; Loftus & Fathi, 1985). Alternatively, the effect could be interpreted more within the framework of social constructivism, with participants being aware that reporting the performer's comments about the bending of the key may make others question the reliability of their testimony, and therefore they omit such comments to increase the perceived accuracy of their observations. This interpretation would be in line with previous work into the way in which some people describe apparently anomalous events that they have experienced, noting the various forms of rhetorical devices (e.g., claiming to be a skeptic prior to experiencing the phenomenon, ruling out all possible normal explanations, etc.) that are used to support the idea that they are a reliable and accurate witness (see, e.g., Woofitt, 1992). Again, it is hoped that these competing interpretations will be assessed in future research. Either way, these findings again underline the notion that testimony relating to certain alleged paranormal pheneomena may not only be an incomplete record of events, but actualy fail to contain the information needed to properly assess potential normal explanations for the alleged affects.

In short, the two studies reported here demonstrate that it is possible to create ostensible PKMB after-effects via verbal suggestion alone. In the second study, participants who reported such effects were more confident in the accuracy of their testimony than others, and tended not to report the performer's suggestions. Taken together these results show the significant impact that a relatively small amount of verbal suggestion can have on the perception and reporting of an ongoing event, and also cast doubt on the notion that eyewitness reports of PKMB after-effects should be seen as evidence in support of the paranormal. It is hoped that other researchers will now use this paradigm to further explore the effects of ongoing verbal suggestion on the perception and reporting of both allegedly anomalous and more commonplace events.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Peter Lamont for helping to make the videotape used in this study, Dr Caroline Watt for her feedback on an earlier version of this paper, and the Perrott-Warrick Fund and Society for Psychical Research for helping to fund this research.

REFERENCES

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1-12.

French, C.C. (1992). Factors underlying belief in the paranormal: Do sheep and goats think differently? *The Psychologist*, *5*, 295-299.

Fuller, U. (1975). *Confessions of a psychic.* Teaneck, NJ: Fulves.

Haraldsson, E. (1985). Interrogative suggestibility and its relationship with personality, perceptual defensiveness and extraordinary beliefs. *Personality and Individual Differences*, *6*, 765-767.

Harris, B. (1985). *Gellerism Revealled.* Calgary: Micky Hades International.

Hasted, J. (1981). *The metal-benders*. London: Routledge & Kegan Paul.

Hergovich, A. (2003). Field dependence, suggestibility and belief in paranormal phenomena. Personality and Individual Differences. 34(2), 195-209

Irwin, H. J. (1993). Belief in the paranormal: A review of the empirical literature. *Journal of the American Society for Psychical Research*, *87*, 1-39.

Irwin, H, (1999). *An Introduction to Parapsychology 3rd ed*. London: McFarland & Co.

Loftus, E.F. (1997). Creating false memories. *Scientific American*, *227(3)*, 70-75.

Loftus, E.F. and Fathi, D.C. (1985). Retrieving multiple autobiographical memories. *Social Cognition*, *3*, 280-95.

Loftus, G.R. (1992). When a lie becomes memory's truth: Memory distortions after exposure to misinformation. *Current Directions in Psychological Science, 1*, 121-123.

Marks, D.F. (2000). *The Psychology of the Psychic (second edition).* Amherst, New York: Prometheus Books.

Marks, D. F. & Kammann, R. (1980). *The Psychology of the Psychic.* Buffalo, NY: Prometheus Books.

Panati, C. (1976). *The Geller Papers.* Houghton Mifflin Company: Boston.

Randi, J. (1975). *The Magic of Uri Geller*. New York: Ballantine.

Randi, J. (1983a). The Project Alpha Experiment: Part 1. The first two years. *Skeptical Inquirer*, *7(4)*, 24-33.

Randi, J. (1983b). The Project Alpha Experiment: Part 2. Beyond the laboratory. *Skeptical Inquirer*, *8(1)*, 36-45.

Roediger, H.L., III, Wheeler, M.A., & Rajaram, S. (1993). Remembering, knowing and reconstructing the past. In D.L. Medin (Ed.), *The Psychology of learning and motivation: Advances in research and theory, vol. 30*. Orlando, FL: Academic Press.

Saucer,P.R., Cahoon, D.D., Edmonds, E. M. (1992). The Paranormal Belief Scale and the Atheistic Ideation Reference Scale as predictors of hypnotic suggestibility. Psychology: A Journal of Human Behavior. 29(1), 44-46.

Truzzi, M. (1987). Reflections on 'Project Alpha': Scientific experiment or conjuor's illusion? *Zetetic Scholar*, *12/13*, 73-98.

Wiseman, R., Greening, E., and Smith, M. (2003). Belief in the paranormal and suggestion in the séance room, *British Journal of Psychology*, *94(3)*, 285-297.

Wiseman, R. & Morris, R. L. (1995a). *Guidelines for Testing Psychic Claimants.* Buffalo, NY: Prometheus Press.

Wiseman, R. & Morris, R.L. (1995b). Recalling pseudo-psychic demonstrations. *British Journal of Psychology*, *86*, 113-125.

Woofitt, R. (1992). *Telling Tales of the Unexpected: The organization of factual discourse.* London: Harvester/Wheatsheaf

Address for correspondence: Richard Wiseman,
Psychology Department, University of Hertfordshire
Hatfield, Herts, AL10 9AB, UK.
E-mail: r.wiseman@herts.ac.uk

MEASURING SUPERSTITIOUS BELIEF: WHY LUCKY CHARMS MATTER

Richard Wiseman[1], & Caroline Watt[2]
[1]University of Hertfordshire
[2]University of Edinburgh

ABSTRACT

A large body of research has attempted to develop theories about the function and origin of superstitious beliefs on the basis of the psychological correlates of such beliefs. Most of this work has measured superstitious belief using the Paranormal Belief Scale (PBS), and has tended to find that superstitious belief is associated with poor psychological adjustment such as low self-efficacy and high trait anxiety. However, the PBS refers solely to negative superstitions (e.g., breaking a mirror will cause bad luck) and omits items referring to positive superstitions (e.g., carrying a lucky charm will bring good luck). Positive superstitions may serve different psychological functions to negative superstitions. Indeed, as with other forms of "positive illusions", beliefs in positive superstitions may be psychologically adaptive. This paper reports two studies investigating this neglected aspect of the psychological correlates of superstitious belief. If positive and negative superstitious beliefs serve different psychological functions, then we might expect, using Analysis of Variance, to find interactions between superstition type, and various relevant individual difference measures. Study 1 was a large-scale Internet-based study which investigated the relationship between endorsement of superstition type, gender, and a single-item measure of neuroticism. Participants were asked to indicate the degree to which they endorsed three negative and three positive superstitious beliefs using five response options (anchored with Definitely Yes and Definitely No). The three negative items were: "Have you avoided walking under a ladder because it is associated with bad luck?"; "Would you be anxious about breaking a mirror because it is thought to cause bad luck?"; and, "Are you superstitious about the number 13?". The three positive items were: "Do you say 'fingers crossed' or actually cross your fingers?"; "Do you say 'touch wood' or actually touch or knock on wood"; and, "Do you sometimes carry a lucky charm or object?". 4,339 individuals took part in study 1. The highly statistically significant results found interactions between gender and endorsement of superstition type, and between neuroticism and superstition type. Study 2 was conducted by mail, and sought to replicate and extend the findings obtained in Study 1 by administering validated questionnaire measures of neuroticism and life satisfaction alongside positive and negative superstition items. 116 individuals took part. As with study 1, there was a significant interaction between gender and superstition type, but no interaction was found for neuroticism. A significant interaction was found between superstition type and life satisfaction. Overall, these findings indicate that the psychological correlates of superstitious belief vary depending on whether the belief is in positive or negative superstitions. These findings have important implications for theory development, demonstrate that the PBS is an incomplete measure of superstitious belief, and highlight the need for future measures to include items referring to positive superstitions.

INTRODUCTION

Paranormal, superstitious and magical beliefs have been found in a diverse range of cultures for thousands of years (Jahoda, 1969), and polls show that these beliefs continue to thrive in modern times (Newport & Strausberg, 2001). Researchers have long speculated about the origins and functions of such beliefs (see, e.g., Frazer, 1922; Jahoda, 1969; Malinowski, 1948; Vyse, 1997). Over the last two decades, almost all of this work has involved attempting to identify reliable psychological, and behavioural, correlates of such beliefs (Irwin, in press). This line of enquiry was stimulated by the publication of the Paranormal Belief Scale (PBS) in the mid-1980s (Tobacyk & Milford, 1983; Tobacyk, 1988). The PBS was the first measure of paranormal belief to be psychometrically evaluated and, although there has been some recent debate over its factorial structure (e.g., Lawrence, 1995; Tobacyk & Thomas, 1997) and internal validity

(Lange, Irwin, & Houran, 2000), it continues to be the most widely-used measure of paranormal belief (Goulding & Parker, 2001).

A significant amount of this research has centered around just one of the seven PBS sub-scales, namely, belief in superstition. A large body of work has demonstrated that, almost without exception, the psychological correlates of this sub-scale reflect relatively poor psychological adjustment, including low self-efficacy (Tobacyk & Shrader, 1991); high trait anxiety (Wolfradt, 1997); irrational beliefs (Roig et al., 1998); an external locus of control (Dag, 1999; Tobacyk, Nagot & Miller, 1988); magical ideation (Thalbourne, Dunbar, & Delin, 1995; Tobacyk & Wilkinson, 1990); psychopathology (Dag, 1999); field dependence and suggestibility (Hergovich, 2003); and dissociative experiences (Wolfradt, 1997). Gender differences have also emerged, with women tending to show higher levels of superstitious beliefs than men (e.g., Dag, 1999; Vyse, 1997; Wolfradt, 1997).

These findings have been used by some researchers as a basis for various theoretical models attempting to explain the prevalence of superstitious thinking, with almost all of this work exploring the role played by such thinking in the initiation and maintenance of maladaptive beliefs and behaviour (e.g., Alcock, 1981; Dag, 1999; Vyse, 1997). For instance, many authors have suggested that paranormal and superstitious beliefs may develop in anxious individuals with a strong need for control, in an attempt to overcome perceived uncertainty in their surroundings (Irwin, 2000; Jahoda, 1969; Malinowski, 1948), or as a coping mechanism following traumatic childhood experiences (French & Kerman, 1996; Irwin, 1992; Lawrence et al., 1995; Ross & Joshi, 1992). This model is supported by recent theoretical developments within cognitive and emotion research, suggesting that anxiety plays a central role in negative emotions (Brown, Chorpita, & Barlow, 1998), and that childhood experiences of diminished control may lead to the development of anxiety (Chorpita & Barlow, 1998).

The superstition sub-scale of the PBS contains three items: ‘Black cats can bring bad luck’; ‘If you break a mirror, you will have bad luck’; and, ‘The number “13” is unlucky’. All of these items refer to beliefs that can be classified as “negative” superstitions – that is, they all reflect the notion that certain behaviours (e.g., breaking a mirror) or omens (e.g., seeing a black cat) are magically associated with unlucky and potentially harmful consequences. Given that this is the case, it is perhaps not surprising that, as noted above, scores on this sub-scale correlate with a range of measures reflecting poor psychological adjustment. However, not all superstitious beliefs fall into this category. Some, such as carrying a charm to bring good luck, touching wood, and crossing fingers, reflect a desire to bring about beneficial consequences by actively courting good luck or at least avoiding bad luck. Such “positive” superstitions may serve different psychological functions to negative superstitions. Indeed, as is the case with other forms of so-called 'positive illusions' (Taylor, 1989), beliefs in these types of superstitions may actually be psychologically adaptive rather than maladaptive. Unfortunately, almost all previous work into the correlates of superstitious belief has used the PBS, and is thus based on an instrument only measuring negative superstitions. As such, this research, and the theoretical work driven by the results of this work, may have failed to explore a vitally important aspect of superstitious thinking. Although a few authors have remarked upon the potential theoretical and practical importance of positive superstitious beliefs (Irwin, in press; Tills, cited in Haining, 1990; Vyse, 1997), we are not aware of any research that has investigated this question empirically. The present paper addresses this issue. Participants were asked to complete several individual difference measures and then indicate the degree to which they endorsed negative and positive superstitions. It was hypothesised that Analysis of Variance (ANOVA) would reveal interactions between superstition type and individual difference measures.

The present paper reports on two studies. Study 1 was a large-scale internet-based study which investigated the relationship between endorsement of superstition type, gender, and a single-item measure of neuroticism. Study 2 sought to replicate and extend the findings obtained in Study 1 by administering validated questionnaire measures of both neuroticism and life satisfaction.

STUDY 1 METHOD

In March 2003, the first author initiated a two-month-long, Internet-based study into superstitious beliefs, as part of an on-going research program into the psychology of luck. The study was promoted through British National Science Week, by articles in broadsheet newspapers which invited members of the public to visit a website and complete a questionnaire. The present analyses examined the resulting database to discover whether the main measures of individual differences used during the study interacted with participants' belief in negative and positive superstitions.

Questionnaire

Participants were first asked to indicate basic demographic information about themselves, including their age (Categories: Under 20; 21-30; 31-40; 41-50; Over 50) and gender. They were then asked to indicate their agreement with a single-item measure of self-perceived neuroticism ("I tend to worry about life") via five response options anchored with Strongly Agree (scoring 5) and Strongly Disagree (scoring 1). Finally, participants were asked to indicate the degree to which they endorsed three negative and three positive superstitious beliefs using five response options (anchored with Definitely Yes and Definitely No). The three negative items concerned walking under a ladder ('Have you avoided walking under a ladder because it is associated with bad luck?), breaking a mirror ('Would you be anxious about breaking a mirror because it is thought to cause bad luck?') and the number 13 ('Are you superstitious about the number 13?'). The three positive items concerned crossing fingers ('Do you say "fingers crossed" or actually cross your fingers?'), touching wood ('Do you say "touch wood" or actually touch or knock on wood?') and carrying a lucky charm ('Do you sometimes carry a lucky charm or object?').

STUDY 1 RESULTS

Participants. 4,339 participants took part in the study, 1951 males and 2388 females. There were 450 individuals in the 'under 20 years' age group, 965 in the '21-30' category, 995 in the '31-40'category, 902 in the '41-50' category, and 1027 in the 'over 50 years' category.

Scoring. Responses to each of the superstition items were transformed into a 5-point scale (ranging from 5 - Definitely Yes, to 1 - Definitely No). The scores relating to the three positive items were summed to provide a measure of the degree to which participants endorsed these superstitions. Likewise, scores relating to the three negative items were summed to provide the same measure for these superstitions. For all participants, the mean neuroticism score was 2.7 (SD = 1.2). To obtain the clearest possible picture of individual differences for the measure of self-perceived neuroticism, only those who responded at the extreme ends of the scale (i.e., responding 'Strongly Agree' or 'Strongly Disagree') were included in the analyses. There were 536 participants in the 'high' self-perceived neuroticism group, and 309 in the 'low' group (see table 1 for further details).

	N	Positive Superstition	Negative Superstition
Gender			
Male	1951	8.0 (3.4)	6.4 (3.6)
Female	2388	10.1 (3.0)	8.0 (3.9)
Self-perceived neuroticism			
High neuroticism	536	10.0 (3.5)	8.7 (4.1)
Low neuroticism	309	7.6 (3.4)	5.2 (3.1)

Table 1: Study 1 descriptive statistics: Number (N) of participants, and mean positive and negative superstition scores (Standard Deviation in parentheses), for each individual difference measure.

Sampling method. The methodological issues surrounding Internet-mediated research (IMR) have been the topic of recent debate within psychology. Some researchers have questioned whether internet-accessed samples are representative of more general populations (see, e.g., Schmidt, 1997), whilst others have argued that IMR usually results in samples that are as representative as those associated with more traditional research paradigms (see, e.g., Hewson, 2003). To help address this issue, the authors examined whether the patterns of belief in negative superstitions within the present data-set matched those reported in previous research. As reported above, past work with the PBS has indicated that women are more superstitious than men, and that people who are anxious are more superstitious than those who are not. The present data-set found the same significant patterns: women exhibited significantly higher belief in negative superstitions than men ($F[1,4337] = 191.31$, $p < .0001$, $eta = .21$); and, the Spearman Rank correlation coefficient between self-perceived neuroticism and belief in negative superstitions for all participants was both positive and significant ($N = 4339$, *Rho* [corrected for ties] = .27, p [2 tailed] < .0001). These patterns provide strong evidence that the data-set obtained in this study was not dissimilar to the data-sets used in this type of research in the past.

Gender. A 2x2 ANOVA (superstition type x gender) revealed a highly significant main effect of gender: women tended to endorse both types of superstition to a greater extent than men ($F[1,4337] = 379.5$, $p < .0001$, $eta = .28$). Overall, positive superstitions were endorsed more than negative superstitions, ($F[1,4337] = 1259.69$, $p < .0001$, $eta = .47$). There was also a highly significant interaction, whereby the difference between the genders narrowed for negative superstitions ($F[1,4337] = 20.40$, $p < .0001$, $eta = .07$).

Self-perceived neuroticism. A 2x2 ANOVA (superstition type x self-perceived neuroticism) revealed a highly significant main effect of neuroticism: high neurotics endorsed both types of superstition more strongly than low neurotics ($F[1,843] = 163.7$, $p < .0001$, $eta = .40$). There was also a highly significant interaction ($F[1,843] = 19.42$, $p < .0001$, $eta = .15$), such that the difference between the groups narrowed for the positive superstitions[1].

STUDY 1 DISCUSSION

Study 1 was a large-scale internet study which investigated whether the psychological correlates of endorsing positive versus negative superstitions might differ. Using analysis of variance, the study examined interactions between belief in positive and negative superstitions, and two measures of individual differences (gender and self-perceived neuroticism). Interactions were found for these individual difference measures, indicating that it is indeed theoretically important for questionnaire measures of superstitious belief to include and differentiate between negative and positive superstitions.

One limitation of study 1 is its use of a single-item indicator of self-perceived neuroticism, which may only have face validity. It was therefore decided to conduct a second study, using a well-established questionnaire measure of neuroticism. Study 1 also was limited in that it took only neuroticism as a measure of psychological adjustment. As noted in the introduction, superstitious belief has traditionally been linked to a wide variety of indicators of poor psychological adjustment. However, in line with the recent surge of popularity in 'positive psychology', we thought it interesting to examine the potential relationship between superstition type and a validated measure of life satisfaction.

[1] It could be argued that people might say “touch wood” or “fingers crossed” through habit and not because they are superstitious. To investigate this issue, we carried out two post-hoc ANOVAs comparing the question relating to lucky charms with the three combined negative superstition items. The results continued to show significant interaction effects for both individual difference measures (Gender $F[1,4337] = 89.98$, $p < .0001$, $eta = .14$; Self-perceived neuroticism $F[1,843] = 106.41$, $p < .0001$, $eta = .33$) strongly suggesting that the original results from the combined measure of positive superstition were not solely due to the “touch wood” and “fingers crossed” items.

STUDY 2 METHOD

Study 2 examined whether the degree to which people endorse different types of superstition (negative versus positive) may vary according to three individual difference measures: gender, neuroticism, and life satisfaction. A volunteer panel built up by the first author through his research into luck, was contacted by email and invited to participate in a postal questionnaire study about superstition and luck. Questionnaires were sent out, and completed questionnaires were returned in postage-paid envelopes.

Questionnaires

The questionnaire pack consisted of three questionnaires. Questionnaire One was the 48-item Revised Eysenck Personality Questionnaire short scale (EPQ-R) (Eysenck & Eysenck, 1991), a well-established measure which includes 12 items designed to indicate Neuroticism, based on Yes/No answers to short statements. High scores indicate high neuroticism. This questionnaire also asked participants' age and gender. Questionnaire Two was the 5-item satisfaction with life scale (Diener et al., 1985), in which participants respond to each statement on a 7-point scale (response options anchored with strongly disagree and strongly agree). High scores indicate high satisfaction with life. Questionnaire Three was the same superstition questionnaire as was used in study 1 (see above for details of content and scoring).

STUDY 2 RESULTS

Participants. 153 questionnaire packs were sent out, and completed questionnaires were returned by 116 participants, giving a 76% response rate. There were 28 male respondents, 77 females, mean age 42 years, SD = 10.12, range 25-66 years. 11 respondents did not give their gender, and are therefore not included in the gender analyses below. Overall mean neuroticism score was 5.7 (SD = 3.8) and overall mean life satisfaction score was 21.8 (SD = 8.5).

Analysis. For the purpose of analysis of variance, participants were split into "high" or "low" neuroticism groups, and "high" or "low" life satisfaction groups, according to a median split for each variable. There were 61 participants in the high neuroticism group (mean = 8.8, SD = 2.1), 55 in the low neuroticism group (mean = 2.2, SD = 1.8), 66 in the high life satisfaction group (mean = 28.1, SD = 3.6), and 50 in the low life satisfaction group (mean = 13.4, SD = 5.2). Mixed ANOVAs were conducted between the two measures of superstitious belief (positive versus negative) and the measures of gender, neuroticism, and life satisfaction (see table 2 for descriptive statistics).

	N	Positive Superstition	Negative Superstition
Gender			
Male	28	7.9 (3.9)	7.0 (3.4)
Female	77	9.6 (3.6)	7.1 (3.6)
Neuroticism			
High neuroticism	61	10.3 (3.4)	8.4 (3.6)
Low neuroticism	55	8.0 (3.8)	5.8 (3.0)
Life satisfaction			
High satisfaction	66	9.2 (3.8)	6.6 (3.5)
Low satisfaction	50	9.3 (3.7)	8.0 (3.6)

Table 2: Study 2 descriptive statistics: Number (N) of participants, and mean positive and negative superstition scores (Standard Deviation in parentheses), for each individual difference measure.

Gender. A 2x2 ANOVA (superstition type x gender) revealed that, as with study 1, women tended to endorse superstitious beliefs more highly than men (see table 2 for details). However, the trend did not reach significance in study 2, perhaps due to relatively low statistical power ($F[1,103] = 1.43$, $p = .23$, *eta* = .12). Also, as before, positive superstitions received higher levels of endorsement than negative superstitions ($F[1,103] = 21.07$, $p < .0001$, *eta* = .41). Study 2 replicated the finding of study 1 of an interaction between superstition type and gender: the difference between men and women in endorsement of positive superstitions narrowed dramatically for negative superstitions ($F[1,103] = 4.83$, $p = .03$, *eta* = .21).

Neuroticism. A 2 x 2 ANOVA (superstition type x neuroticism) found a significant main effect for neuroticism ($F[1,114] = 18.86$, $p < .0001$, *eta* = .38). This replicated the finding in study 1 that high neurotics tend to endorse both types of superstitious beliefs more highly than low neurotics. However, unlike study 1, study 2 found no interaction between superstition type and neuroticism ($F[1,114] = .31$, $p = .58$, *eta* = .05)[2]. The overall correlation between neuroticism and belief in negative superstitions in the present data-set again supported the findings from previous research ($N = 116$, *Rho* [corrected for ties] = .39, *p* [2 tailed] < .0001). The Spearman Rank correlation coefficient between the N and L scales on the EPQ-R was non-significant, suggesting that the results were not unduly influenced by dissimulation ($N = 116$, *Rho* [corrected for ties] = -.08, *p* [2 tailed] =.38).

Life satisfaction. A 2 x 2 ANOVA (superstition type x life satisfaction) found a tendency for individuals low in life satisfaction to endorse superstitions more highly, however the trend was not significant ($F[1,114] = 1.55$, $p = .22$, eta = .12). A significant interaction was found, whereby the difference between the life satisfaction groups dramatically reduced for positive superstitions ($F[1,114] = 4.30$, $p = .04$, eta = .19).

GENERAL CONCLUSIONS

The vast majority of research examining the psychological correlates of superstitious belief has used the superstition sub-scale of the PBS. This sub-scale contains three items, all of which refer to negative superstitions. We present two studies examining potential difference in patterns of belief for positive and negative superstitions. Significant interactions were found, for four out of five analyses, with effect sizes ranging from .07 to .21. These findings raise questions about the validity of previous research and have theoretical and methodological implications for future research.

On a theoretical level, these results have important implications for those wishing to understand why people hold superstitious beliefs. As noted towards the start of this paper, almost all of the theoretical work in this area has viewed superstitious thinking within the context of the initiation and maintenance of maladaptive beliefs and behaviour. The significant interactions found in the present studies underline the importance of expanding this theoretical understanding to take account of both positive and negative superstitions. The required expansion maybe relatively small and easily incorporated into traditional models associating superstitious belief with psychological maladjustment. This would be the case if, for example, future research revealed that the mechanisms underlying belief in positive superstitions are conceptually similar to those that drive belief in negative superstitions. Alternatively, belief in positive superstitions could have their basis in quite different mechanisms, such as the promotion of self-efficacy and optimism, and thus may only be fully explained via theoretical approaches that are substantially different to existing models.

[2] Median split was chosen for this analysis in study 2 in order to maximise statistical power. To enable further comparison with study 1, which used extreme scorers on the neuroticism measure, a post hoc analysis was conducted for study 2 taking approximately the top and bottom 25% of neuroticism scorers. There were 32 in the "low" neuroticism group (mean N = 0.9, SD = 0.8) and 35 in the "high" neuroticism group (mean N = 10.4, SD = 1.2). Analysis of Variance showed no significant interaction between neuroticism and superstition type ($F[1,65] = .06$, $p = .80$, *eta* = .03). Therefore this post hoc analysis also did not replicate the interaction found in study 1.

On a methodological level, these findings strongly suggest that it is important that any valid measure of superstitious belief includes reference to both positive and negative superstitions. The frequently-used PBS superstition sub-scale fails to do this, and thus there is a pressing need for the patterns found in previous studies using the scale to be interpreted as correlates of belief in negative superstitions, rather than superstition *per se*.

Future research should aim to develop a broader measure of belief that encompasses much wider, and much more diverse, forms of superstitions. There is clearly a need for a more fine-grained understanding of the psychological functions of different superstition types, beginning with the fundamental positive versus negative distinction that we have highlighted in this paper.

ACKNOWLEDGEMENTS

The authors are grateful to Professor Ian Deary, Professor Harvey Irwin, and Professor Robert Morris, for their helpful comments on an earlier draft of this manuscript. This research was supported by the Perrott-Warrick Fund and the Koestler Parapsychology Unit.

REFERENCES

Alcock, J. E. (1981). *Parapsychology: Science or Magic? A Psychological Perspective*. Elmsford, NY: Pergamon Press.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107*, 179-192.

Chorpita, B. F., & Barlow, D. H. (1998) The development of anxiety: The role of control in the early environment. *Psychological Bulletin, 124*, 3-21.

Dag, I. (1999). The relationships among paranormal beliefs, locus of control and psychopathology in a Turkish college sample. *Personality and Individual Differences, 26*, 723-737.

Diener, E., Emmons, R.A., Larson, R.J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71-75.

Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult)*. London: Hodder & Stoughton.

Frazer, J. G. (1922). *The Golden Bough*. London: Macmillan.

French, C. C., & Kerman, M. K. (1996). *Childhood trauma, fantasy proneness and belief in the paranormal*. Paper presented at the Annual Conference of the British Psychological Society, London.

Goulding, A., & Parker, A. (2001). Finding psi in the paranormal: Psychometric measures used in research in paranormal beliefs/experiences and in research on psi-ability. *European Journal of Parapsychology, 16*, 73-101.

Hergovich, A. (2003). Field dependence, suggestibility and belief in paranormal phenomena. *Personality and Individual Differences, 34*, 195-209.

Hewson, C. (2003). Conducting research on the Internet. *The Psychologist, 16*, 290-293.

Irwin, H. J. (1992). Origins and functions of paranormal belief: The role of childhood trauma and interpersonal control. *Journal of the American Society for Psychical Research, 86*, 199-208.

Irwin, H. J. (2000). Belief in the paranormal and a sense of control over life. *European Journal of Parapsychology, 15*, 68-78.

Irwin, H. J. (in press). *The Psychology of Paranormal Belief. Parapsychological Monographs No. 20.* New York: Parapsychology Foundation.

Jahoda, G. (1969). *The Psychology of Superstition*. Harmondsworth, England: Penguin.

Lange, R., Irwin, H. J., & Houran, J. (2000). Top-down purification of Tobacyk's revised Paranormal Belief Scale. *Personality and Individual Differences, 29*, 131-156.

Lawrence, T. R. (1995). How many factors of paranormal belief are there? A critique of the Paranormal Belief Scale. *Journal of Parapsychology, 59*, 3-25.

Lawrence, T., Edwards, C., Barraclough, N., Church, S., and Hetherington, F. (1995). Modelling childhood causes of paranormal beliefs and experiences: Childhood trauma and childhood fantasy. *Personality and Individual Differences, 19*, 209-215.

Malinowski, B. (1948). *Magic, Science and Religion*. Garden City, NY: Doubleday.

Newport, F., & Strausberg, M. (2001). *Americans' belief in psychic and paranormal phenomena is up over last decade*. Princeton: Gallup News Service.

Roig, M., Bridges, K. R., Renner, C. H., & Jackson, C. R. (1998). Belief in the paranormal and its association with irrational thinking controlled for context effects. *Personality and Individual Differences, 24*, 229-236.

Ross, C. A., & Joshi, S. (1992). Paranormal experiences in the general population. *Journal of Nervous and Mental Disease, 180*, 357-361.

Schmidt, W. C. (1997). World Wide Web survey research: Benefits, potential problems and solutions. *Behaviour Research Methods, Instruments and Computers, 29*, 270-273.

Taylor, S. E. (1989). *Positive Illusions: Creative Self-Deception and the Healthy Mind*. New York: Basic Books.

Thalbourne, M. A., Dunbar, K. A., & Delin, P. S. (1995). An investigation into correlates of belief in the paranormal. *Journal of the American Society for Psychical Resarch, 89*, 215-231.

Tills, H., cited in Haining, P. (1990). *Superstitions*. London: Treasure Press.

Tobacyk, J. (1988). *A Revised Paranormal Belief Scale*. Unpublished manuscript. Louisiana Tech University, Rushton, LA.

Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology, 44*, 1029-1037.

Tobacyk, J., Nagot, E., & Miller, M. (1988). Paranormal beliefs and locus of control: A multidimensional examination. *Journal of Personality Assessment, 52*, 241-246.

Tobacyk, J., & Shrader, D. (1991). Superstition and self-efficacy. *Psychological Reports, 68*, 1387-1388.

Tobacyk, J., & Thomas, A. (1997). How the big orthogonal seven is really the oblique seven. *Journal of Parapsychology, 61*, 337-342.

Tobacyk, J., & Wilkinson, L. V. (1990). Magical thinking and paranormal beliefs. *Journal of Social Behavior and Personality, 5*, 255-264.

Vyse, S. A. (1997). *Believing in Magic: The Psychology of Superstition*. New York: Oxford University Press.

Wolfradt, U. (1997). Dissociative experiences, trait anxiety and paranormal beliefs. *Personality and Individual Differences, 23*, 15-19.

Address for correspondence: Caroline Watt,
University of Edinburgh, Koestler Parapsychology Unit,
7, George Square, Edinburgh EH8 9JZ, UK.
E-mail: caroline.watt@ed.ac.uk

TESTING ALLEGED MEDIUMSHIP: METHODS AND RESULTS

Richard Wiseman[1] & Ciarán O'Keeffe[2]
[1]*Psychology Department, University of Hertfordshire*
[2]*Psychology Department, Liverpool Hope University*

ABSTRACT

Mediums claim to be able to communicate with the deceased. Such claims attract a considerable amount of public interest and, if valid, have important implications for many areas of psychology and parapsychology. For over one hundred years, researchers have tested alleged mediums. This work has obtained mixed results and provoked a considerable amount of methodological debate, with some researchers arguing that some mediums possess genuine psi abilities and others arguing that such alleged abilities are the result of psychological, rather than parapsychological, processes. This paper reviews the key issues in this debate, noting that this past methodological debate has tended to cluster around three important issues, namely; the need to control for sensory leakage, the need to assess the generality of mediums' statements and the need for blind judging. The paper then describes how the authors devised a method of testing that prevented the problems that have hindered past research and how they then used this method to test several professional mediums. This test involved five mediums giving readings to five sitters. For all of the readings, the sitters were located in a different room to the mediums, and neither the mediums, nor anyone who came in contact with the mediums, knew the identity of the sitters. In addition, the time of each reading was counter-balanced across several days in order to minimize the possibility of any temporal cues. The resulting readings were then broken up into separate statements and given to the sitters for evaluation. The sitters were not informed whether the statement had been made during their own reading or the reading of another sitter. The ratings assigned by each of the sitters to each of the readings were then assessed using via a form of Monte-Carlo analysis. The results of this work did not support the existence of genuine mediumistic ability as none of the mediums obtained significant results. Competing interpretations of these results are discussed, including the notion that the previous work reporting significant results may be methodologically flawed, that the mediums employed in the present study did not possess psi abilities and that this study was carried out under conditions that were not psi conducive. The authors conclude by urging other researchers to adopt the methodology reported here to investigate other mediums and psychics, and also explore ways in which the methodology presented in the paper could be used to assess conceptually similar, but non-paranormal, claims made in clinical, occupational and forensic contexts.

INTRODUCTION

Some individuals claim to possess mediumistic abilities that allow them to contact the 'spirit world' and receive information from the deceased. There are several reasons to subject these claims to rigorous and empirical investigations.

First, mediumistic abilities, if valid, would provide evidence to support the survival of bodily death, and thus have important implications for aspects of psychology and parapsychology. Such data would, for example, present a strong challenge to key assumptions underlying neuropsychological research, including the notion that human personality, cognition and consciousness is dependent on a living brain. Evidence of genuine mediumistic abilities would also raise intriguing questions about the sensory mechanisms that might underlie such abilities and, on a more practical level, have important implications for the many aspects of clinical and counselling psychology concerned with bereavement and grief.

Second, demonstrations of apparent mediumistic abilities have a significant impact on public belief and behaviour. Recent opinion polls have revealed that almost 30% of Americans now believe in the existence of genuine mediumistic abilities (Newport and Strausberg, 2001), approximately 10% of Britains visit mediums to both receive messages from the deceased and obtain general guidance for their lives (Roe, 1998), and new types of television programmes featuring such demonstrations consistently attract millions of viewers (Brown, 2001). Well-controlled tests of mediums would help the public and television programme makers assess the validity of such alleged abilities, and thus help inform their resulting decisions and behaviour.

Third, certain individuals working in non-paranormal contexts make claims that are analogous to those made by mediums, and the methods developed to test mediums could be used to examine these claims. For example, some clinicians claim to be able to gain insights into patients' backgrounds purely from their reactions to certain projective tests, some practitioners working in an occupational setting appear to be able to give detailed accounts of people's personality simply from their scores on certain assessment tools, and some individuals operating in a forensic context claim to be able to produce accurate profiles of offenders from a very limited amount of behavioural information. Several writers (e.g., Wood, Nezworski, Lilienfeld, Garb, 2003: Alison, Smith and Morgan, 2003) have recently noted that the anecdotal evidence supporting these claims may be the result of the same types of psychological stratagems that can underlie the apparent accuracy of mediumistic readings (i.e., the use of general statements, chance, etc.), and thus the methods developed to examine such claims may benefit from a thorough understanding of the procedures used to test mediumiship.

Given the profound nature of the theoretical and practical issues surrounding this topic, it is perhaps not surprising that the scientific testing of mediumship has a long and controversial history.

Initial tests of mediums were carried out in the 1880s and primarily involved investigators attending séances, noting down the comments that the mediums allegedly received from the deceased, and then attempting to assess the accuracy of this information. The majority of the resulting reports argued in favour of the existence of genuine mediumistic ability, and contained lengthy transcripts of mediumistic messages along with detailed descriptions of the evidence supporting these statements (see, e.g., Hodgson, 1892, 1898). Critics have attacked this work, arguing that it often failed to assess whether the seemingly accurate readings could have been the result of various psychological stratagems, such as the mediums engaging in shrewd guesswork or producing very general statements that would be endorsed by the majority of people (see, e.g., Podmore, 1901; Hyman, 1977; Gardner, 1992).

Over the years, several researchers have attempted to devise procedures that eliminate the potential for such stratagems, and then used these to examine some of the best-known mediums of the day. The resulting studies have obtained mixed results, with some work finding evidence in favour of genuine paranormal abilities and other research supporting the null hypothesis (for a review of this work, see Schouten, 1994). This work has provoked a considerable amount of methodological and statistical debate, much of which has focused on the degree to which the procedures employed in those studies obtaining positive results have eliminated potential biases and problems (for a recent example of this type of debate, see Hyman, 2002; Schwartz, 2003, Hyman 2003). Unfortunately, discussion surrounding the diverse range of potential methodological and statistical issues that can bias such work is spread across the psychological and parapsychological literature, and is often presented in a piecemeal way rather than being more conceptually organised. In addition, researchers working in this area have yet to develop a relatively standard method of testing that is both practical and minimises the potential for such artefacts. This paper addresses both of these issues. The first part of this paper reviews the main problems that have hindered previous tests of mediumship and describes how the authors devised a method of testing that was both practical and methodologically sound. The second part of the paper presents a detailed description of how this method was then used to test several professional mediums.

METHODOLOGICAL ISSUES: PROBLEMS AND PROCEDURES

The debate concerning the potential problems that can arise during tests of alleged mediumistic ability has centred around three key issues; (i) the need to control for potential sensory leakage, (ii) the need to accurately assess the generality of the mediums' statements and (iii) the need for 'blind' judging. The following three sections briefly review each of these problems and outlines the types of procedures that can be employed to overcome them.

The need to control for potential sensory leakage

Alleged mediums may be able to gain information about their clients (often referred to as 'sitters') via normal means and then use this information to help produce accurate readings. Such information may be obtained in a variety of ways. For example, books about how to fake mediumstic abilities describe various techniques for obtaining useful information in advance of a reading, including, for example, secretly eaves-dropping on sitters' conversations, or conducting surreptitious searches of telephone directories and the Internet (see, e.g., Roland, 1998). Other writers have described how experienced mediums may be able to unconsciously gain information from more subtle sources, such as sitters' clothing, posture, demeanour and jewellery (see, e.g., Morris, 1986). Even a very limited amount of contact between medium and sitter has the potential to provide useful information. For example, Wiseman and O'Keeffe. (2001) noted that the speed with which the sitter answers 'yes' or 'no' to the medium's questions could unconsciously provide experienced mediums with useful feedback about the accuracy of their comments during a reading.

For these reasons, any well-controlled test of mediumistic ability should prevent mediums gaining information about sitters via normal means. This usually involves experimenters taking appropriate steps to ensure that mediums cannot ascertain any information about sitters in advance of test readings, and that there are sufficient safeguards preventing them from obtaining verbal or nonverbal cues from them during the readings. Such safeguards should also extend to anyone involved in the study (e.g., experimenters or other participants) who are aware of any information about the identity of the sitters. Researchers examining the possible existence of telepathy have developed various procedures for eliminating potential sensory leakage between participants (see, e.g., Milton & Wiseman, 1997) and many of these safeguards (e.g., placing participants in separate rooms with sufficient levels of sound attenuation) can be employed to eliminate possible leakage during tests of alleged mediumistic abilities.

The need to assess the generality of mediums' statements

Research into the so-called 'Barnum Effect' has consistently shown that people tend to rate certain types of very general personality statements (e.g., 'You have a great deal of untapped creative potential') as highly accurate (Forer, 1949; Furnham & Schofield, 1987). In addition, more recent work has revealed that even statements that do not appear especially general can be true of many people. For example, Blackmore (1994) carried out a large-scale survey in which over 6000 people were asked to state whether quite specific statements were true of them. Over one third of people endorsed the statement 'I have a scar on my left knee' and over a quarter answered yes to the statement 'Someone in my family is called Jack'. Mediums can utilise this phenomena to produce readings that may appear highly accurate but, in reality, simply contain very general statements that are endorsed by a large number of sitters.

Attempts to deal with this issue in tests of mediumistic ability have taken many forms over the years and have been the subject of considerable debate. In perhaps the earliest attempt to solve the problem, Hyslop (1919) collated statements that had been endorsed by a sitter during various test readings, and then asked a 'control' group containing approximately 500 people to indicate whether each statement was true of them. Hyslop then calculated the general acceptance level of the reading on basis of the percentage of people in the control group that endorsed each statement. For example, if 250 people in the control group endorsed the statement 'you are male', then Hyslop calculated the probability of acceptance as 250/500 or 0.5. To obtain an overall probability of all of the statements being endorsed, Hyslop multiplied the individual probabilities for each of the statements together (e.g., the probability of two statements being endorsed, each having a general acceptance level of 0.5, would be .25). Several critics have correctly noted that this approach greatly inflates the medium's apparent accuracy because it incorrectly assumes that each of the statements are independent of one another (Schouten, 1994). Thus, if, for example, the medium stated that the sitter 'had recently lost someone who was male' and that this person 'had a beard', the probability of these statements would be multiplied together as if they were independent, whereas the probability of the first being accurate is heavily related to the probability of the second being correct.

Over the years, researchers have devised various forms of analyses that attempt to overcome this problem (see, for example, Saltmarsh & Soal, 1930, Pratt, 1936). Probably the most widely endorsed and employed is that developed by Pratt and Birge (1948). In the Pratt and Birge procedure, a small number of sitters each receive a reading from a medium. The sitters are then asked to rate the accuracy of statements from both their own reading (often referred to as the 'target' reading) and those from the readings of other sitters (referred to as 'decoy' readings). If the medium is accurate, then the ratings assigned to the target readings will be significantly greater than those assigned to the decoy readings. If, however, the medium is simply producing general statements, then the sitters will assign similar ratings to both the target and decoy readings. Pratt and Birge noted that the results of experiments using this procedure can perhaps best be viewed as shown in Table 1, with the numbers on the diagonal of the table (shown in bold) representing the scores that each sitter gave to their own readings, and the numbers on the off-diagonal numbers representing the ratings that sitters assigned to the readings of others (Pratt, 1969).

		Sitter judging accuracy of reading				
		John	Eric	Bill	Tony	Tom
Sitter present during reading	John	**58**	23	46	6	56
	Eric	25	**73**	14	45	53
	Bill	18	41	**67**	33	39
	Tony	61	22	40	**49**	30
	Tom	11	39	26	28	**72**

Table 1: Standard way of representing data from experiments employing the Pratt and Birge technique.

It is widely recognised that the statistical analyses used to test whether the numbers on the diagonal are significantly greater than those on the off-diagonals do not assume that the statements within the readings are independent (Pratt, 1969). To this end, researchers have recommended using Monte-Carlo analyses that create a distribution of the sum of the numbers on the diagonal for each possible permutation of the matrix, and then calculate the probability of the experimental outcome by examining where the sum of the numbers on the diagonal actually obtained in the experiment lies within this distribution (for further discussion about such analyses see, Pratt & Birge, 1948; Thouless, 1949; Greville, 1949; Pratt, 1969; Scott, 1972).

The need for 'blind' judging

The way in which sitters rate the accuracy of mediumistic readings is highly subjective (Hyman, 1977). For example, Wiseman and O'Keeffe (2001) note that the statement 'The spirits are talking about the younger woman who has now passed away', is open to several interpretations (e.g., the word 'younger' could refer a young child, a teenager, or even someone who died in their forties), and that the degree to which a sitter is prepared to think through these alternative interpretations will influence the perceived accuracy of the statement. The process of assessment can also be biased by selective recall. For example, the medium saying 'Your daughter was an extrovert' may cause sitters to selectively recall certain life events (i.e., the times that his or her daughter went to parties), forget other events (e.g., the times that she wanted to be alone), and thus assign a spuriously high accuracy rating to the statement. The degree to which the sitter thinks about alternative interpretations of ambiguous statements and engages in selective recall may be influenced by several factors, including, for example, their need to believe in the afterlife or please the medium.

Researchers testing alleged mediumistic ability have attempted to eliminate such biases by having sitters rate the accuracy of statements without informing them whether the statements are drawn from target or decoy readings (Pratt, 1969). However, such procedures may not fully eliminate some of the more subtle temporal cues that might help sitters distinguish target from decoy readings. Imagine, for example, that the sittings are scheduled for different days and that in one reading the medium refers to a memorable news story (e.g., 'The spirits are upset by that horrible train crash today'). When the sitters are subsequently presented with the readings for assessment, they may see this comment and correctly deduce the day on which the reading took place, and thus know whether that this is their target reading. Similar problems may arise even if the sitters are scheduled at different times on the same day, if the medium's comments allow a sitter to figure out when a reading was made (e.g., during a lunchtime sitting the mediums remarks, 'The spirits always get hungry around now'), or both the sitter and medium experience an idiosyncratic event during a reading (e.g., a crash of lightning outside) and the medium makes reference to this event (e.g., 'The lightning is making it difficult to contact the spirits').

To our knowledge, previous tests of alleged mediumistic ability have failed to recognise, and therefore control for, this potential artefact (although see Milton and Wiseman, 1997 for a discussion regarding how the same type of temporal cues could bias the outcome of certain types of extra-sensory perception experiments). Various procedures could be employed to minimise the problem. For example, the sitters can be scheduled on the same days, and the time of their readings can be counter-balanced across the days (i.e., each sitter has one session scheduled at 11.30 a.m', one at 12.30 a.m. etc.). Also, the sitter and/or the medium can be located in room that isolate them obvious sources of idiosyncratic external events, such as unusual weather conditions or noise from surrounding rooms and corridors. Finally, the statements that make-up the reading can be separated and randomly ordered before being presented to sitters for assessment, thus minimising the possibility of a subtle cue in one statement influencing the way in which sitters evaluate an entire reading. None of these procedures will fully eliminate the potential problems associated with temporal cuing. For example, if the medium were to say a statement that contained clues about both the day and time it was produced (e.g., "I simply cannot get the images of yesterdays terrible train crash out of my mind – I think it happened around about now"), then sitters may be able to figure out whether this statement was made during a target or decoy reading. However, the suggested procedures will help eliminate most of the main forms of cuing, and it seems unlikely that a medium would produce a large number of the type of statements described above.

AN EXPERIMENTAL TEST OF MEDIUMSHIP

The previous section outlined the main methodological and statistical problems that can hinder tests of mediumship, and some of the procedures that can be employed to eliminate these potential problems. As noted above, many researchers carrying out such tests have discussed, and attempted to control for, these potential artifacts. For example, Pratt (1969) outlined the need to eliminate any sensory leakage between medium and sitters, assess the generality of statements, and have sitters judge the accuracy of both target and decoy readings without knowing which statements were produced during their own readings. However, some previous work in this area has not employed safeguards against some of these potential problems (see, e.g., Schwartz et al., 2001) and, to our knowledge, no previous testing of mediumship has controlled for some of the more subtle forms of bias discussed above (e.g., the need to eliminate potential temporal cues that might help sitters distinguish target from decoy readings). The authors recently devised a method for testing mediumship that incorporated all of the procedures described above, and then used this method to test several professional mediums. This sections outlines the methods and results of that test.

The test involved five professional mediums giving readings for five sitters under conditions that eliminated any potential sensory leakage between medium and sitter. The sitters were then asked to assess the accuracy of the mediums' statements without knowing whether the statements were drawn from target or decoy readings. Monte-Carlo analyses were then used to assess whether the ratings assigned to target readings were significantly higher than the ratings assigned to decoy ratings.

Participants

Mediums: The 5 mediums (3 female, 2 male; age range 42-55) were recruited via a list of certified mediums provided by the Spiritualists Nationalist Union (SNU). The SNU stated that all of the mediums on this list had undergone a rigorous selection procedure and were subject to continual assessment. Each medium was initially contacted by telephone, and then sent a detailed description of the protocol and consent form.

Sitters: The 5 sitters (all male, ages range: 25 – 30) were either students or staff from the University. They were selected from a pool of individuals who responded to a general email, circulated within the University, asking for volunteers to be involved in a scientific test of mediumship. The sitters were chosen using the following criteria; (i) they did not know one another, (ii) they were the same gender and (iii) they were approximately the same age. Each sitter was initially contacted by telephone, and then sent a detailed description of the protocol and consent form. None of the sitters were paid for their involvement in the study.

Rooms and apparatus

The experiment took place in a suite of rooms located within the University's Psychology Department (see Figure 1). The medium was located in the studio area and the sitter was placed in the meeting room. These rooms were acoustically isolated from one another, such that the sitter could not hear the medium and vice versa. Events happening outside the building (weather effects, etc.) could not be heard in the meeting room and noise from the corridor directly outside the meeting room could not be heard in the studio. The mediums' comments were recorded via video cameras operated remotely by E1 from the control room. The sitter was supplied with a portable stereo system and headphones, in order that they could listen to music throughout the session.

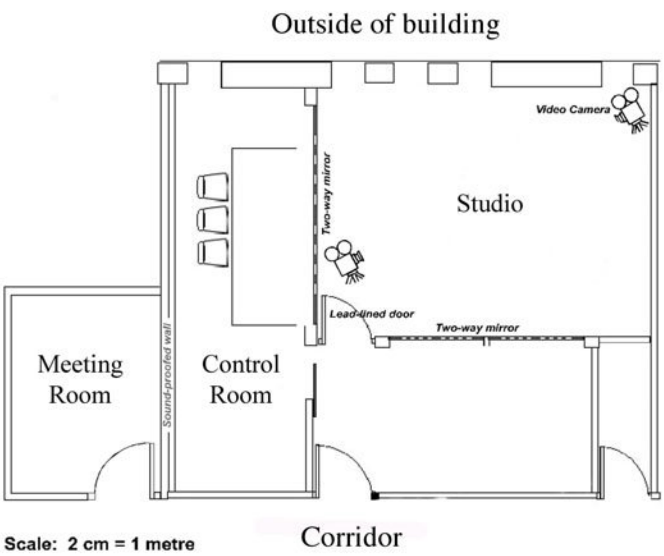


Figure 1: Floor-plan of rooms

Procedure

The experiment was run by two experimenters, E1 and E2. E1 initially selected and contacted the five mediums, whilst E2 selected and contacted the five sitters. E1 then liased with the mediums to arrange a day on which each of them could visit the University to participate in the study (referred to as 'experimental days'). E1 then passed these five dates to E2, who arranged for each sitter to visit the Psychology Department at a specified time on each of these days. E2 scheduled the five sitters in hourly slots from 11.00 a.m. to 4.30 p.m.. The order of the sitters was counterbalanced across the five days (see Table 2).

	Mediums				
	M1	M2	M3	M4	M5
11.00 - 12.00	A	E	D	C	B
12.00 - 13.00	B	A	E	D	C
13.30 - 14.30	C	B	A	E	D
14.30 - 15.30	D	C	B	A	E
15.30 - 16.30	E	D	C	B	A

Table 2: Counter-balanced order of sitters A-E on each experimental day

At the start of each experimental day, E1 met the medium and took him or her to the studio. E1 outlined the nature and design of the study, and ensured that the medium was comfortable. At 11.00 a.m., E1 informed the medium that the first session was just about to start and then left the studio for the control room. The medium was allowed to say as much or little as he or she wished during the next sixty minutes. All of these comments were recorded on videotape and monitored by E1 from the control room. At approximately 12.00 midday, E1 re-entered the studio and informed the medium that the session had ended, and that the next session would commence after a short break. This procedure was repeated five times throughout the day.

E2 met the first sitter at 10.45 a.m. and took them to the meeting room. They were given a choice of music and asked to listen to this music, via the headphones provided, between 11.00 a.m. and midday. At approximately 11.00 a.m., E2 left the meeting room and remained outside until the end of the session. At approximately midday E2 re-entered the meeting room, terminated the session and thanked the sitter for participating. This procedure was repeated five times throughout the day.

In the event of any unexpected events (e.g., a session time having to change slightly due to the late arrival of a sitter), the two experimenters communicated with one another via a series of simple coded text messages.

Throughout the study, both E1 and the mediums did not receive any information about the identity of the sitters, E2 did not have any contact with the mediums, and all of the mediums and sitters remained blind to each others' identities.

After the five experimental days had been completed, E1 transcribed the mediums' comments, removed any extraneous details from the transcripts (e.g. questions to the experimenter, pauses) and broke each reading down into a series of statements (see Appendix A for example). The statements from all of the mediums were then randomly mixed together and sent to the sitters, who were asked to rate the accuracy of each statement between 1 (Not applicable) and 7 (Very applicable). Each sitter independently rated all of the statements, and an overall score for each reading was created by summing the individual ratings assigned to each of the statements making up that reading.

RESULTS

The ratings were collated for each medium into a Pratt and Birge table (see Appendix B), and the significance levels for each dataset calculated via a Monte-Carlo analysis based on the 120 possible permutations of each matrix.

None of the analyses were significant and the resulting p-values (1 tailed) were as follows; Medium 1: 0.89; Medium 2: 0.27: Medium 3: 0.27: Medium 4: 0.77: Medium 5: 0.66: All mediums combined: 0.70. Inspection of the data revealed that there was only one occasion (medium 2, reading for sitter B) when the sitter for whom a reading was intended assigned a higher rating to the reading than the other four sitters. On all other occasions, the ratings assigned by sitters who were not present at the time of the reading were higher than the rating assigned by the sitter for whom the reading was intended.

DISCUSSION

This paper briefly first outlined the major methodological and statistical problems that have hindered previous test of alleged mediumistic abilities, and described procedures that can be used to minimise these problems. It then described the way in which these procedures were implemented during the authors' recent test of five professional mediums. This test involved five sitters each receiving five readings and then rating the accuracy of those readings. The results revealed that the ratings that sitters assigned to their own readings were not significantly different from the ratings they assigned to others sitters' readings, and thus did not support the existence of mediumistic ability.

These findings can be interpreted in various ways. It is possible that genuine mediumstic ability does not exist, and that the apparent accuracy of mediums' readings are entirely due to the type of psychological stratagems outlined in the first section of this paper. This interpretation is consistent with much of the sceptical literature on alleged mediumship (see, e.g., Hyman, 1977: Gardner, 1992), previous work that has also failed to find evidence of such abilities under controlled conditions (see Schouten, 1994 for a review of this work) and those arguing that the studies that have obtained positive results are methodologically flawed (e.g., Hyman, 2002, 2003). If this interpretation is correct, then the most productive direction for future work in this area is to examine these stratagems more closely, examining, for example, the types of people that tend to endorse mediumistic readings and the forms of rhetoric that alleged mediums use to convince sitters that they are receiving messages from their deceased friends and relatives. Obviously, it is not possible to conclude that any phenomenon does not exist on the basis of a single study. However, additional studies producing similar results, using a diverse range of mediums and sound methodology, would add weight to this interpretation. Alternatively it is possible that genuine mediumistic abilities do exist, but that this study failed to find evidence of them because, for example, the mediums involved in the experiment do not possess such abilities or the setting in which the study was conducted did not elicit such abilities. These hypotheses can only be evaluated by systematically varying these factors in future work, providing that such work also eliminates the various methodological problems discussed in this paper. For example, many of the studies into mediumship that have obtained highly positive results were conducted around the turn of the last century, and used trance mediums, rather than individuals who claim to communicate with the deceased whilst in a waking state. Future work could explore this hypothesis by recruiting mediums who appear to be in an altered states of consciousness when receiving spirit messages.

On a methodological level, the study eliminated the various types of bias that can hinder research in this area. Whilst some of these procedures have been used in several previous tests of mediumship (e.g., safeguards against sensory leakage between mediums and sitters, and use of the Pratt-Birge technique), others have not been utilised in this context before (e.g., safeguards against potential temporal cues). The resulting methodology was both practical and straightforward, and it is hoped that other researchers will employ this method to investigate other individuals claiming similar types of paranormal abilities, and that this work will help tease apart the competing interpretations outlined above.

Finally, as noted in the Introduction, certain individuals working in clinical, occupational and forensic contexts make claims that are analogous to those made by mediums (i.e., being able to ascertain highly accurate information about a person or situation on the basis of very limited data), and thus tests of such claims could benefit from many of the methodological procedures described in this paper. For example, evaluations of the efficacy of the types of projective tests widely used within certain areas of clinical psychology should involve sufficient safeguards against sensory leakage between clinician and participant, take into account the potential generality of statements produced by the clinician and ensure that those statements are 'blind' judged in such a way as to minimise possible temporal cues. Up to this point in time, the literature discussing the potential problems that can hinder research attempting to assess mediumistic claims, and the possible procedures that can be employed to overcome them, has been widely distributed across a range of highly specialist publications within parapsychology. It is hoped that the conceptual grouping of these problems and procedures presented in this paper, along with an example of how they were combined into what the authors believe to be a practical and methodologically sound method that was used to assess several professional mediums, will help bring this work to psychologists working in a broad range of applied contexts.

In short, the present study found no evidence to support the notion that the professional mediums involved in the research were, under controlled conditions, able to demonstrate paranormal or mediumistic ability. However, the authors believe that they have developed a practical, straightforward and methodologically sound way of testing such claims, and it is hoped that this approach will be employed by researchers to test other individuals who appear to have mediumistic or psychic abilities, and conceptually similar claims being made within clinical, occupational and forensic contexts.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Caroline Watt for her feedback on an earlier version of this paper, the mediums and sitters who kindly participated in the study, and the Perrott-Warrick Fund for helping to fund the research.

REFERENCES

Alison, L. J., Smith, M. D. & Morgan, K. (2003). Interpreting the accuracy of offender profiles. *Psychology, Crime and Law, 9,* 185-195.

Blackmore, S. (1994). Probability Misjudgement and belief in the paranormal: Is the theory all wrong? *In D. Bierman (Ed.), Proceedings of the 37th Annual Convention of the Parapsychological Association.* 72-82.

Brown, Ivy. (2001). Hearing from dearly departed proves a hit on Sci-Fi Channel. *Los Angeles Times*, March 5.

Forer, B. R. (1949): The fallacy of personal validation: a classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology*, 44, 118-123.

Furnham, A., Schofield, S. (1987): Accepting personality test feedback: a review of the Barnum Effect. *Current Psychological Research and Reviews,* 6, 162-178.

Gardner, M. (1992). *On the wild side*. Buffalo, NY: Prometheus Books.

Greville, T.N.E. (1949). On the number of sets required for testing the significance of verbal material. Journal of Parapsychology, 13, 137-138.

Hodgson, R. (1892). A record of certain phenomena of trance. *Proceedings of the Society for Psychical Research*, 8, 1-167.

Hodgson, R. (1898). A further record of observations of certain phenomena of trance. *Proceedings of the Society for Psychical Research*, 13, 284-582.

Hyman, R. (1977). Cold reading: How to convince strangers that you know all about them. *The Skeptical Inquirer*, 1, 18-37.

Hyman, R. (2002). How Not to Test Mediums: Critiquing the Afterlife Experiments. *Skeptical Inquirer*, 26(1).

Hyman, R. (2003). Hyman's Reply to Schwartz's 'How Not To Review Mediumship Research', *Skeptical Inquirer*, 27(3), 61-64.

Hyslop, J.H. (1919). Chance coincidence and guessing in a mediumistic experiment. *Proceedings of the American Society for Psychical Research*, 13, 5-88.

Milton, J. & Wiseman, R. (1997). *Guidelines for extrasensory perception research.* University of Hertfordshire Press: Hatfield, England.

Morris, R. (1986). What psi is not: the necessity for experiments. In H.T. Edge, R.L. Morris, J. Palmer, J.H. Rush (Eds). *Foundations of parapsychology*. London: Routledge & Kegan Paul.

Newport, F. and Strausberg, M. (2001). Americans' Belief in Psychic and Paranormal Phenomena Is up Over Last Decade. Report by The Gallup Oranisation, June 8, 2001.

Podmore, F. (1901). On Professor Hyslop's report on his sittings with Mrs. Piper. *Proceedings of Society for Psychical Research, 17,* 374-388.

Pratt, J.G. (1936). Towards a method of evaluating mediumistic material. *Bulletin of the Boston Society for Psychic Research*, 23, 5-53.

Pratt, J.G. & Birge, W.R. (1948). Appraising verbal test material in parapsychology, *Journal of Parapsychology*, 12, 236-256.

Pratt, J. G. (1969). *On the Evaluation of Verbal Material in Parapsychology* (Parapsychology Monograph No 10). New York: Parapsychology Foundation.

Roe, C. A. (1998). Belief in the Paranormal and Attendance at Psychic Readings. *Journal of the American Society for Psychical Research, 92*(1), 25-51.

Rowland, I. (1998). *The Full Facts Book of Cold Reading*. Ian Roland Limited: London, England.

Saltmarsh, H.F. & Soal, S.G. (1930). A method of estimating the supernormal content of mediumistic communications. *Proceedings of the Society for Psychical Research*, 39, 266-271.

Schouten, S.A. (1994). An Overview of Quantitatively Evaluated Studies With Mediums and Psychics. *The Journal of the American Society for Psychical Research*, 88, 221-254.

Scott, C. (1972). On the evaluation of verbal material in parapsychology: A discussion of Dr Pratt's monograph. *Journal of the Society for Psychical Research, 46*, 752, 79-90.

Schwartz, G.E.R., Russek, L.G.S., Nelson, L.A. & Barentsen, C. (2001). Accuracy and Replicability of Anomalous After-Death Communication Across Highly Skilled Mediums. *Journal of the Society for Psychical* Research, 65(862), 1-25.

Schwartz, G.E. (2003). How *Not* To Review Mediumship Research, *Skeptical Inquirer*, 27(3), 58-61.

Thouless, R.H. (1949). Review of Journal of Parapsychology, Vol 12, Number 4. *Journal of the Society for Psychical Research, 35*, 48-49.

Wiseman, R., and O'Keeffe, C. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A critique. *The Paranormal Review*, 19: 3-6.

Wood, J.M., Nezworski, M.T., Lilienfeld, S.O., Garb, H.N. (2003). The Rorschach Inkblot Test, Fortune Tellers, and Cold Reading. *Skeptical Inquirer, 27*(4), 29-33.

Address for correspondence: Richard Wiseman,
Psychology Department, University of Hertfordshire
Hatfield, Herts, AL10 9AB, UK.
E-mail: r.wiseman@herts.ac.uk

APPENDIX A

EXAMPLE OF A READING BEING CONVERTED INTO CORRESPONDING STATEMENTS

Reading

With this particular reading, I was sensing a beard and I was sensing a room where the walls go up and curve inwards, so more of an ornate type room. I would say there is a link with a different culture with this one, probably Islam or Muslim link. The light in the room is a fairly normal one, nothing fancy or flash so its just a normal lamp with a normal sort of lampshade so even though its an ornate type of room, its not really a posh room that this person is in. Now again all I'm feeling with the ear, possibly someone who's got an earring in the ear, one of those modern ones that people wear, in the left ear this is. Now with the person I'm giving the reading to, I sense that there's something that needs to be cut off, they need to cut off from something, something they're holding onto that they need to let go of.

Statements

With this particular reading, I was sensing a beard.

I was sensing a room where the walls go up and curve inwards, so more of an ornate type room.

I would say there is a link with a different culture with this one, probably Islam or Muslim link.

The light in the room is a fairly normal one, nothing fancy or flash so its just a normal lamp with a normal sort of lampshade

Now again all I'm feeling with the ear, possibly someone who's got an earring in the ear, one of those modern ones that people wear, in the left ear this is.

Now with the person I'm giving the reading to, I sense that there's something that needs to be cut off, they need to cut off from something, something they're holding onto that they need to let go of.

APPENDIX B

RESULTS TABLES FOR INDIVIDUAL MEDIUMS AND ALL MEDIUMS COMBINED.

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**156**	230	131	176	252
	B	202	**183**	223	192	301
	C	170	111	**145**	106	161
	D	348	240	196	**166**	175
	E	120	97	134	159	**137**

Results for Medium 1

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**354**	410	304	294	322
	B	198	**277**	210	243	230
	C	234	289	**282**	318	264
	D	452	341	220	**269**	113
	E	166	250	275	291	**258**

Results for Medium 2

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**15**	21	13	7	12
	B	11	**17**	9	24	12
	C	13	13	**12**	15	7
	D	8	26	18	**24**	11
	E	14	13	10	21	**10**

Results for Medium 3

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**72**	42	98	43	109
	B	51	**59**	49	67	89
	C	88	61	**77**	110	146
	D	97	84	54	**61**	91
	E	122	57	64	84	**87**

Results for Medium 4

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**257**	318	410	194	181
	B	410	**364**	312	286	243
	C	355	289	**297**	321	176
	D	312	320	354	**331**	238
	E	427	276	358	246	**220**

Results for Medium 5

		Sitter judging accuracy of reading				
		A	B	C	D	E
Sitter present during reading	A	**854**	1021	956	714	876
	B	872	**900**	803	812	875
	C	860	763	**813**	870	754
	D	1217	1011	842	**851**	628
	E	849	693	841	801	**712**

Results for all mediums combined.

PANEL: PERSPECTIVES IN THE STUDY OF MEDIUMSHIP

CHAIR: CARLOS S. ALVARADO

Carlos S. Alvarado, Ruth Reinsel, Peter Mulacz, Rosemarie Pilkington, Erlendur Haraldsson

ON TRANCES AND SPIRIT COMMUNICATORS: RESEARCH IDEAS FROM THE OLD LITERATURE

Carlos S. Alvarado
University of Virginia

There are many aspects of mediumship we can study from the psychological perspective. This includes consideration of cognitive, psychosocial, and psychopathological variables. I will discuss the topic using suggestions and observations from the old psychical research literature.

To this day we do not understand the nature and functions of trance. There is much to do with the stages and depth of trance. Cesare Lombroso (1909) argued that the most striking physical phenomena with Eusapia Palladino took place in the deepest trance. Eleanor Sidgwick (1915) documented the existence of stages of trance and the variety of phenomena accompanying Mrs. Piper's mentation. Just as modern parapsychologists have studied ESP in relation to aspects of altered states of consciousness (e.g., depth, changes in sense of time and body image) we may study psi in mediums in relation to specific features of trance.

There is also much to do regarding the imagery obtained in mediumistic mentation. We may attempt to replicate James H. Hyslop's (e.g., 1919) studies of the role of visual imagery, and its distortions, on the generation of veridical and non-veridical mentation. There are probably many individual differences in the content of the mediumistic mentation. But regardless if the mentation shows common patterns or not it would be worthwhile to explore the flow of imagery, its symbolism, and salient features such as the repeated use of specific images in veridical messages, as documented by Charles Drayton Thomas (1939) in his study of Gladys Osborne Leonard.

Also important is the psychosocial study of the content of mediumistic statements and the development of trance personalities. Williams James (1890) emphasized that the Zeitgeist was influential on mediumistic productions. The writings and studies of Théodore Flournoy (1900, 1911), Pierre Lebiedzinski (1924), and Réne Sudre (1926) show the importance of indirect suggestions, beliefs and demand characteristics on the development of trance personalities and the stories of the communicators. Most of the previous discussions have been theoretical and after the fact, but there is much to do to induce or manipulate specific ideas to explore the plasticity of the medium's psychological resources. These manipulations could involve the production of psi effects such as communications from the living. Positive results would place ESP explanations of mediumship on empirical footing.

We also need to return to the old question of the relationship of mediumship to psychopathology. In the old days the psychopathological nature of mediumship was defended both by those skeptical of its psi components, such as Pierre Janet (1889), and those that believed in psi aspects such as physical phenomena (e.g., Morselli, 1908). Some mediums seem to show unstable behaviors or a tendency towards instability. Perhaps we should study the issue considering Joseph Maxwell's (1903/1905) suggestion that the nervous system of mediums is liable to many changes and fluctuations that do not necessary become pathological. This is consistent with some contemporary ideas of schizotypy that postulate that some people are more creative or hallucinatory than others due to differences in the inhibitory mechanisms of the nervous system. Assuming this is the case with mediums we should expect to find more adjustment and productive use of mediumistic phenomena in some than in others, according to different developmental and situational variables.

Address for correspondence: Carlos S. Alvarado, University of Virginia Health System, Division of Personality Studies, P. O. Box 800152, Charlottesville, VA 22908.
E-mail: alvarado@parapsychology.org

PHYSICAL MEDIUMSHIP, MACRO-PK AND THE AUTONOMIC NERVOUS SYSTEM

Ruth Reinsel
The NeuroPsience Laboratory, Hillsdale

The accounts of séances held with several physical mediums of the 19th and early 20th century provide tantalizing clues to extreme bodily states during the production of PK effects. Observations of mediums such as Eusapia Palladino, Franek Kluski, and Rudi Schneider give evidence that macro-PK phenomena were produced at a substantial physical cost to the medium.

Symptoms that were commonly observed during macro-PK included pallor, increased heart rate (tachycardia), hyperventilation or irregular breathing (dyspnea), hoarseness of voice, and increased perspiration. After séances where macro-PK phenomena were apparently produced without fraud, many physical mediums experienced severe physical and mental sequelae, including in some cases, exhaustion, headache, extreme muscle weakness, vomiting, and long-lasting disturbances of digestive function and sleep.

These symptoms from the séance room share a common basis in that they are produced by competing branches of the autonomic nervous system (ANS) which exert opposing effects on bodily organs. The sympathetic branch has a stimulatory effect and produces arousal, while the parasympathetic branch produces relaxation. Normally both branches are active at the same time, to keep the body in an optimum state of functioning.

Braud (1981, 1985) reviewed the evidence that the autonomic nervous system may modulate psi performance. He concluded that GESP is generally facilitated by calm, quiet states, and PK may be facilitated by states of arousal. Roll and Persinger's (1998) observations with poltergeist agents support the role of arousal in macro-PK. The symptoms listed above during production of macro-PK are symptoms of activation of the sympathetic branch of the ANS. The post-séance sequelae are typical of activation of the parasympathetic branch of the ANS (Robertson, Low & Polinsky, 1996).

Most of the cardiovascular and respiratory effects and alterations in level of arousal seen during physical mediumship are under the control of the vagus nerve (cranial nerve X) and the insular cortex. The insular cortex (insula) lies directly under the temporal lobe. It controls the regulation of bodily functions by the ANS. Sympathetic activation of the insula results in increased heart rate and blood pressure, respiration, salivation, and dilation of the pupil, as well as gastrointestinal activity (Robertson, Low & Polinsky, 1996).

There seems to be substantial overlap between these functions of insular cortex and the responses of physical mediums during PK production.

It is postulated here that at the end of a prolonged sympathetic "storm" associated with the production of macro-PK phenomena, a compensatory rebound in parasympathetic activity occurs. This sudden decrease in sympathetic activity and increase in parasympathetic activity may cause the symptoms of muscle weakness, headache, dizziness, nausea and fainting (syncope, in medical terminology).

If macro-PK is indeed produced by hyperstimulation of sympathetic neurons in the ANS, controlled by insular cortex and brainstem regions, then it is likely to be a relatively primitive function from an evolutionary standpoint. Since, in this view, the control centers are not located in the cerebral cortex, PK (and perhaps ESP?) is likely to be not under conscious control; it would be nonverbal, and characterized by visceral sensations ("gut feelings" and vague "malaise" or formless anxiety). This description fits well with what we know about the functioning of psi. We may need to look beyond (beneath?) the cerebral cortex to find the brain centers that facilitate macro-PK.

Address for correspondence: Ruth Reinsel,
The NeuroPsience Laboratory, 90 Werimus Avenue,
Hillsdale, N. J. 07642, USA.
E-mail: ruthrein@mindspring.com

PHYSICAL MEDIUMSHIP: OPEN QUESTIONS FROM THE PAST — A CHALLENGE FOR THE FUTURE

Peter Mulacz
Austrian Society for Parapsychology

The Rhinean paradigm has proved extremely successful and has gained scientific respectability for the field of parapsychology. Not only was its approach very simple – an experiment that can be replicated at any place and any time using average people without the need to search for specifically gifted ones – moreover this kind of replicability echoed the demand of replicability in natural sciences and thus paved the way to the acceptance of the field in academia.

Yet there is a reverse side to this success story. Investigating specifically gifted individuals (sensitives, mediums) got out of the focus of 'scientific parapsychology' and much of what has been achieved by earlier research fell into oblivion. (Admittedly, for the last 30 years there has been a certain upsurge in the research on gifted individuals in the context of the CIA-sponsored research in RV.) Nonetheless, in particular, research in physical mediumism came out of fashion and was at best left to amateurs. Taking into account the potential relevance of (ostensible) phenomena of physical mediumism the present state of affairs is a deplorable one. However, research in Paranormal Metal Bending was a first little step in the right direction of reversing that process.

The difficulties in investigating physical mediumism are twofold: firstly finding gifted subjects and keeping them motivated, and secondly applying appropriate research methodology. What must be demanded from an appropriate methodology is again twofold: automatic registration of data and safety against mediumistic fraud. Galileo at the very onset of natural science some 400 years ago is credited with saying 'to measure what is measurable, and what is not, to make it measurable', and same still applies to research in physical mediumism. Visual observation (a prime example is the account of Princess Metternich on her observations on D. D. Home), preferably by several witnesses, is good, however, automatic continuous recording of quantitative data is better. This is the standard since the 1870ies (Crookes' experiments with D. D. Home, recording of the readings of a spring balance) and there is no way of going below this level. It should be added that automatic recording of data as such is not necessarily fraud-proof and proper safeguard needs to be taken in any case.

The gold standard of experiments in physical mediumism is beyond any doubt the series of experiments with Rudi Schneider carried out by Eugène and Marcel Osty in 1930. Not only was the methodology of recording the absorption of an infrared beam flawless (at least has it not been called into doubt by fellow parapsychologists), also the results that have been achieved were impressive: the correlation of the oscillation of said absorption rate to the respiration rate of the medium. The nature of this correlation remains an open question (one of the many in the field). The later experiments with Rudi Schneider by G. A. Schwaiger employing an early TV-set were discontinued due to the turn of events during the 2nd World War.

Every leading edge of scientific research needs to utilize state-of-the-art equipment. At the time of Schrenck-Notzing this has been Amereller's electrical medium control apparatus applied in experiments carried out in the dark or in extremely reduced light conditions. The equivalent present level would be video recording such experiments by using light amplifier devices. Any experiment falling short of this standard (e. g. the Scole Group) cannot be regarded as conclusive.

Address for correspondence: Peter Mulacz,
Österreichische Gesellschaft für Parapsychologie und
Grenzbereiche der Wissenschaften,
c/o Institut für Ethnologie, Kultur- und
Sozialanthropologie der Universität Wien,
Universitätsstraße 7, A-1010 Wien.
E-mail: peter.mulacz@aon.at

THE MYSTERY OF ECTOPLASM

Rosemarie Pilkington
Staten Island, New York

Most people consider the production of ectoplasm by physical mediums fraudulent. Many charlatans took advantage of the darkness of the séance room and the desire of believers to communicate with the dead. Many were unmasked and physical mediumship was ever after suspect.

However there are still instances that seem genuine. My purpose here is to reexamine some of the evidence for these cases and to encourage researchers to not "throw out the baby with the bathwater," but to look into a phenomenon that might teach us much about human mind/body capabilities.

Ectoplasm, if it exists, by its very nature strains the credulity of even those whose "boggle thresholds" have been stretched nearly to infinity. It is supposedly made up of organic elements, especially albumin, yet it appears in various forms from a vapor or clear liquid to a plastic, white solid or semi solid. It is exuded from various orifices of the medium's body but can dissipate and/or be reabsorbed by the medium. In addition it is also subject to the psychological idiosyncrasies of individual personalities, and may at times be influenced by what Jung called the "joker" or "trickster" aspect of our psyches.

Its many forms are as varied as the beliefs and unconscious mental images of the mediums who produce it. The phantom "Katie King," produced by Florence Cook, was at times life-like, seemingly warm blooded, had a measurable pulse, spoke and interacted with observers but did not always appear fully formed and disintegrated before the eyes of witnesses under the controlled conditions of her chief investigator, Sir William Crookes.

Ectoplasm may also have electrical or photo-electrical properties and sometimes appears as lights or luminous animate objects. These were particularly pronounced in the emanations of D. D. Home, Franek Kluski, and recently the Scole mediums.

The phenomena have been widespread with credible investigators reporting on cases throughout Europe and America beginning in the nineteenth century, and continuing throughout the twentieth.

Many photographs have been taken of ectoplasmic forms. These would be unconvincing except for the accompanying information provided by investigators and corroboration by witnesses.

Of interest are the findings of W. J. Crawford, a lecturer in mechanical engineering who examined the mediumistic work of Kathleen Goligher in the early 1900s. Among his theories and discoveries were ectoplasmic "rods" that seemed to cause the levitation of objects. He also found relationships between physical manifestations, including raps and levitations, and the weight, not only of the medium, but of the other participants in the séances.

Crawford was also able to trace the origins and workings of the ectoplasm. His engineering model however did not explain other related phenomena such as luminosity and the seeming suspension of the gravitational field e.g., the adhering of objects to a tilted table or the tilting of a candle's flame when the candle was at an angle. However, his experience may illustrate the concept of ideoplasty in which the experimenter's theory is confirmed because his beliefs affect the medium who in turn shapes the phenomena.

Physical mediums, although out of fashion, still exist. I encourage seeking out those who might lend themselves to investigation. I also want to urge investigators to attempt séances as the "Philip" group in Canada did.

Address for correspondence: Rosemarie Pilkington,
30 Donna Court 5, Staten Island,
NY 10314, USA.
E-mail: rpilkin603@aol.com

A SURVEY OF MEDIUMS IN ICELAND

Erlendur Haraldsson
University of Iceland

We contacted and interviewed all persons we were able to find with experiences of trance or automatisms, that is, who had experienced involuntary speech or writing, at least a few times. These were 31 persons, 17 men and 14 women ranging in age from 19 to 70 (M = 45). 18 of them had fallen in trance more than 100 times, so most of them were practicing mediums, either professionally or in private home circles or organised groups like healing groups. About half of them admitted receiving pay for their psychic work and for two it was this their main income.

These mediums could be roughly divided into three rather equal-sized groups,

a) Spiritists, where the emphasis was on communication with the deceased.
b) Groups seeking healing from spirit helpers and deceased doctors who continued their healing work after death.
c) Group exclusively found in Iceland, "nyalssinnar", who believe that after death we are reborn on other planets in other solar systems, and can then communicate with people on this earth either through dreams or trance.
d) The main purpose of this survey was to learn about how the mediumship started and developed in these individuals. About half of them first fell into trance at home (four of them being alone), or in the homes of friends and did so unexpectedly and involuntary, wheras 40% experienced mediumship as a result of trying to fall into trance or after training. It also happened that the first trance took place unexpectedly and involuntary at a seance when someone else was in trance.

About 41% were fully conscious when they experienced their first trance, which then consisted of involuntary speech where they felt that that some entity had taken control of their organs of speech. Another 40% knew nothing about what happened and had full amnesia about it, and were told about the incident by those who had been present. The rest, 20%, were inbetween, being partially conscious and observing what happened, and partially unaware of what took place.

77% felt that a distinct personality took over when they felt into their first trance. In most instances this personality was unknown to the individual, was from distant times, or a foreign person, like an Indian, or a contemporary person generally known (like a physician) but had not been known to the individual. With time the relationship with this distinct otherworldly personality became quite close but we had the impression that with time little further knowledge was obtained about the life of that personality who became the mediums control.

Very few experienced the trance as an unpleasant, mostly it was experienced as neutral or pleasant. All our interviewees looked upon these personalities as deceased persons, none of them considered them a fiction of their mind or a part of their unconscious. For them these were external beings distinct from themselves.

Address for correspondence: Erlendur Haraldsson,
Faculty of Social Science, University of Iceland,
Oddi v/Sturlugötu, 101 Reykjavik, Iceland.
E-mail: erlendur@hi.is

PANEL: COUNSELING PERSONS WITH EXCEPTIONAL EXPERIENCES: THE EXAMPLE OF IGPP

CHAIR: EBERHARD BAUER

Eberhard Bauer, Martina Belz-Merk, Wolfgang Fach, Niko Kohls & Harald Walach

INTRODUCTION: COUNSELING AND MENTAL HYGIENE – THE APPROACH OF THE IGPP

Eberhard Bauer
Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg

Since its foundation in the year 1950 by Prof. Hans Bender (1907-1991), the pioneer of German post-war parapsychology, the IGPP has provided the public with an information and counseling service covering the whole spectrum of parapsychology, fringe sciences and anomalistic phenomena. Bender, who had a training in psychology as well as in medicine, had always stressed the necessity that research into paranormal phenomena should also consider psychological and social problems connected with popular forms of spiritism or occultism. Already back in the 1950s, Bender had published several case studies which dealt with what he called spiritistic or mediumistic psychoses (e.g., Bender 1955). Therefore, according to Bender's legacy, the IGPP had always to offer information and psychological help for people who have to cope with occult or paranormal experiences causing them and other emotional distress. In a more general sense, "Mental Hygiene" (or "Psycho Hygiene"), as defined the bylaws of Institute, comprises "the application of medical, psychological and parapsychological findings with regard to diagnostics, counseling, intervention and prevention in connection with those scientific questions and psychosocial problems which derive from or are related to anomalous and/or paranormal phenomena." In order to realize this ambitious programme and after new substantial funds became available, the IGPP started since 1996 in collaboration with the Institute of Psychology of Freiburg University a special research project "Counseling and Help for People Claiming Exceptional Experiences" which was directed between 1998 and 2001 by Dr. Martina Belz-Merk. The goal of the project was to develop, implement and evaluate a special counseling and treatment concept for people who felt distressed or burdened by exceptional experiences (EE). In accordance with current regulations and standards of basic documentation in psychotherapy, a special documentation system ("DOKU") was developed by the research group to record systematically sociodemographic, anamnestic and phenomenon-specific data. Using this documentary system, it was possible for the first time to record continuosly and systematically, as well as evaluate statistically, the number of IGPP counseling cases. According to an on-going process of data collection and evaluation, the DOKU System was modified and optimized several times (Belz-Merk, 2002). There exists now a carefully documented data base of counseling cases which can be used for different research strategies (see the panel contributions by Martina Belz-Merk and Wolfgang Fach). In October 2003, the IGPP team organized for the first time a special seminar on practical and theoretical aspects of counseling persons with EE whose audience were primarily medical and psychological therapists and counselors. The IGPP counseling team consists of trained psychologists with a clinical-therapeutical background; the counselors meet each other on a weekly basis to discuss incoming cases. For clients psychotherapeutic treatment can be offered.

Although the necessity for a specific information and counseling service in the field of parapsychology has been increasingly recognized in the last twenty years, especially among the representatives of scientific parapsychology, as exemplified by the PA community (Solvin, 1995) and other parapsychological organisations, professional counseling services dealing with such experiences are still extremely rare. In Holland, Wim Kramer, trained as a clinical psychologist and parapsychologist, had run in Utrecht between 1986 and 1991 in cooperation with the now defunct Parapsychology Laboratory of Utrecht University his pioneering "Parapsychologisch Adviesburo" (Parapsychological Counseling Center) (for his approach see Kramer, 1993). In 1989, the New York based Parapsychology Foundation had organized in London an international conference on "Psi and Clinical Practice" (later published by Coly & McMahon, 1993). In the same year, Dr. Walter v. Lucadou founded in Freiburg, Germany, his "Parapsychologische Beratungsstelle" (Parapsychological Counseling Center) which was later funded by the state of Baden-Wuerttemberg and became well-known in Germany (for his approach see Lucadou & Poser, 1997). In Argentina, the "Instituto de Psicología Paranormal", run by Alejandro Parra, has organized special discussion groups for people with conflictive psi experiences (Montanelli & Parra, 2000). In 2000, as another promising sign, the American Psychological Association (APA) had published the volume *Varieties of Anomalous Experience* (edited by Cardeña et al.) which tried to integrate psychological and parapsychological findings related to exceptional human experiences into the larger corpus of psychology and behavioural sciences.

References

Belz-Merk, M. (2002). *Beratung und Hilfe für Menschen mit Außergewöhnlichen Erfahrungen* [Counseling and Help for Persons with Exceptional Experiences]. Final Report, IGPP.

Bender, H. (1955). Okkultismus als seelische Gefahr [Occultism as a danger for mental health], in M. Pfister-Ammende (ed.), *Geistige Hygiene: Forschung und Praxis [Mental Hygiene: Research and Practice]* (pp. 489-499). Basel: Benno Schwabe.

Cardeña, E., Lynn, S. J. & Krippner, S. (eds.) (2000). *Varieties of Anomalous Experiences*. Washington, DC: American Psychological Association.

Coly, L. & McMahon, J. D. S. (ed.) (1993). *Psi and Clinical Practice*. New York: Parapsychology Foundation.

Kramer, W. H. (1993). Recent Experiences with Psi Counseling in Holland, in L. Coly & J. D. S. McMahon (eds.), *Psi and Clinical Practice* (pp. 124-144). New York: Parapsychology Foundation.

Lucadou, W. v. & Poser, M. (1997). *Geister sind auch nur Menschen* [Ghosts Are Just Humans]. Freiburg: Herder.

Gómez Montanelli, D. E. & Parra, A. (2000). Conflictive psi experiences: A survey with implications for clinical parapsychology, in F. Steinkamp (ed.), *The Parapsychological Association 43th Annual Convention Proceedings of the Presented Papers* (pp. 178-191).Freiburg, Germany.

Solfvin, J. (convenor) (1995). Clinical Parapsychology: A Panel Discussion, in N. J. Zingrone (ed.), *The Parapsychological Association 38th Annual Convention Proceedings of the Presented Papers* (pp. 461-467).

Address for correspondence: Eberhard Bauer,
IGPP, Wilhelmstr. 3a, D-79098 Freiburg, Germany.
E-mail: bauer@igpp.de

THE ROLE OF INCONSISTENCY AND INSTABILITY FOR THE ORIGIN OF EXCEPTIONAL EXPERIENCES

Martina Belz-Merk

Department of Clinical Psychology (University of Freiburg) & Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg

Results of numerous surveys show that paranormal phenomena and Exceptional Experiences (EE) belong to a common human body of knowledge (Bauer & Schetsche, 2003). So it is adequate to assume that for the understanding of the cognitive-emotional functioning of people with EE basically psychological models are relevant.

Paranormal phenomena may represent the surface structure of a broad spectrum of EE. Being able to describe these phenomena may be a necessary but not a sufficient precondition to understand the underlying intrapsychic and interpersonal processes related to the origin and development of EE.

What we know from the empirical literature as well as from our own data (i.e. results of N=858 advice-seeking persons reporting EE to the IGPP) is, that EE are often preceded by a state of substantial instability based on special life events, traumatic experiences or phases of change in life. From a neuronal network point of view this is a central factor for new experiences. The basic hypotheses we are testing are the following: Whenever people experience exceptional phenomena they are and remain in a very tense and instable state for a longer time than would be expected. Usually people fall from one state with strong tensions directly into another attractor. People with EE seem to differ from others just with respect to this point. As a reason we see their special ability to remain in a dissociative state of consciousness (Irwin, 1992) and their special style of information processing (Kuhl & Kazen, 1997).

In order to elaborate a better understanding of the dynamics involved in the structure and psychological functioning of people claiming EE, the so-called Plan Analysis (Caspar, 1995) was used. Plan Analysis is a method, which serves to analyse and describe conscious and unconscious instrumental strategies starting from the level of concrete behaviour up to superordinate general needs. Different aspects of psychological functioning like behaviour, emotion, cognitive schemata but also motivational conflicts can be clearly arranged and reflected as far as their significance in the dynamics is concerned. Then the generated typical plan structure is used as a basis for neural network (connectionist) computer simulations of these individuals. Network models can represent and illustrate the described phenomena and help to shed light on the underlying mechanisms. Simulations which are based on plan-analysis from intensive interviews with N=10 selected cases representing different patterns of EE, additional questionnaire data (SCL90; PSSI: Kuhl & Kazen, 1997; FAMOS: Grosse-Holtforth & Grawe, 2000) will be shown. They model EE from a reactive and instrumental point of view.

The use of such models and its implications for research, counseling and therapy will be demonstrated and discussed.

References

Bauer, E. & Schetsche, M. (eds.) (2003). *Alltägliche Wunder* [*Everydaylife Wonders*]. Würzburg: Ergon.

Caspar, F. (1995). *Plan Analysis. Toward Optimizing Psychotherapy.* Seattle: Hogrefe-Huber.

Grosse-Holtforth, M. & Grawe, K. (2000). Fragebogen zur Analyse motivationaler Schemata (FAMOS). *Zeitschrift für Klinische Psychologie und Psychotherapie, 29*, 170-179.

Irwin, H.J. (1992). Origins and functions of paranormal belief: The role of childhood trauma and interpersonal control. *Journal of the American Society for Psychical Research, 86 (3)*, 199-208.

Kuhl, J. & Kazén, M. (1997). *Das Persönlichkeits-Stil-und-Störungs-Inventar (PSSI): Manual.* Göttingen: Hogrefe.

Address for correspondence: Martina Belz-Merk,
Psychologisches Institut der Universität Freiburg,
Abteilung für Rehabilitationspsychologie,
Engelbergerstraße 41, D-79085 Freiburg, Germany.
E-mail: belz@psychologie.uni-freiburg.de

THE PHENOMENOLOGY OF EXCEPTIONAL EXPERIENCES: AN ANALYSIS OF IGPP COUNSELING CASES

Wolfgang Fach
Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg

There is a great variety of exceptional experiences (EE). Different kinds of specific phenomena often occur jointly and build up to a complex pattern. The purpose of this analysis was to explore if there exist systematic correlations between specific categories of such phenomena. Based on a sample of N = 652 clients who had contacted between 1996 and 2001 the counseling services at the IGPP, six basic patterns of EE could be extracted by using a principal component analysis:

1. *Internal phenomena/feelings of being influenced:* Somatic phenomena and/or voices are experienced. The phenomena are localized physically or inner-psychic, and are often attributed to influences originating from external powers (e.g. by magic, spirits or ghosts).
2. *Apparitions/altered states of consciousness:* Visible apparitions (e.g. light, shadows, figures) and/ or the presence of invisible entities, powers or beings (ghosts, deceased) are experienced in the external world. The phenomena sometimes happen together with sleep paralysis and altered states of consciousness described in terms of "trance" or as "mystic".
3. *Poltergeist phenomena:* Inexplicable kinetic phenomena (objects move, appear or disappear) and/or acoustic sounds (e.g. knocking, taps) are heard in the environment.
4. *Extrasensory perception (ESP):* Past or present events (clairvoyance) are known without usual sensory information transmission, unpredictable future events (precognition), cognitions and emotions of other persons and living beings are known or foreseen.
5. *Mediumship:* Experiences connected with occult and spiritistic techniques such as automatic writing, glass sliding or channeling are experienced as contacts with external powers (ghosts, deceased).
6. *Precognitive dreams:* Dreams in which future events, unpredictable by existing knowledge, are more or less distinctly or symbolically anticipated.

Using these EE patterns specific types of clients can be characterized. A cluster analysis based on the Ward technique yielded six clusters:

1. *Internally influenced type:* 114 persons exclusively reported internal phenomena/influence, the first EE pattern listed above.
2. *Apparition type:* 146 persons reported apparitions/altered states of consciousness. About 50% of them were additionally affected by internal phenomena/influence, more than 40% reported poltergeist phenomena.
3. *Poltergeist type:* 102 persons with poltergeist phenomena. Approximately 30% of them described internal phenomena/influences in addition.
4. *ESP type:* 162 persons concerning ESP. More than 30% of them were also affected by internal phenomena/influences, almost 30% were affected by apparitions/altered states of consciousness, and more than 20% by poltergeist phenomena.

5. *Precognitive dream type*: Typical for 84 persons were experiences of precognitive dreams. 40% of them also reported ESP, almost 30% apparitions/altered states of consciousness, and about 25% poltergeist phenomena and internal phenomena/influence.
6. *Mediumistic type*: The smallest group with 44 persons has exceptional experiences in the context of occult techniques and mediumship. 20% to 30% of them report internal phenomena/influence, poltergeist phenomena and ESP.

The obtained types of clients show significant differences with respect to sex, age, employment status, general stress, specific stress due to EE, experience with psychiatric treatment and assessment of psychological disruption. Future work has to clarify if the obtained typology could serve as a starting point for specific counselling and treatment concepts.

Address for correspondence: Wolfgang Fach,
IGPP, Wilhelmstr. 3a, D-79098 Freiburg, Germany.
E-mail: fach@igpp.de

EXCEPTIONAL EXPERIENCES AND MENTAL HEALTH – RESULTS AND PROBLEMS OF A QUESTIONNAIRE STUDY

Niko Kohls & Harald Walach

University Hospital Freiburg,Department of Evaluation Research in Complementary Medicine at the Institute of Environmental Medicine, in cooperation with the Institut für Grenzgebiete der Psychologie und Psychohygiene (IGPP), Freiburg

Clinical psychology and psychiatry have in general regarded exceptional human experiences as psychopathological in nature, interpreting them mostly as destabilising events. During the last three decades this point of view slowly began to change under the influence of striking research results within the field of salutogenesis and transpersonal psychology: it was discovered that certain types of exceptional experiences can have positive effects on the mental and physical health of individuals. However, the relevant research results, which are mainly based on questionnaire research, are in summary quite heterogeneous, and as a result of this the question arises whether quantitative research methods are appropriate and valid research methods for the scientific investigation of exceptional experiences and their after-effects.

To investigate this unsolved question the Department of Evaluation Research in Complementary Medicine (director: Harald Walach) at the Institute of Environmental Medicine (University Hospital Freiburg), conducted a study ("Freiburg survey of exceptional experiences"), supported by the IGPP, in which we combined quantitative and qualitative research strategies. In this study we included a clinical sample (N = 56), a sample of practitioners of spiritual or religious techniques (N = 350), and a sample of people who don't practice spiritual techniques (N = 299).

In the quantitative method part of the study we used a questionnaire developed by ourselves, which has already been revised, for the measurement of exceptional experiences beside standard instruments for the measurement of sense of coherence, social support, mental distress and transpersonal confidence.

A factor analysis (principal component analysis with varimax rotation) clearly assigned the 57 items of our scale to four factors, accounting for a total of 48% of variance explained. The first factor contains positive mystical experiences (e.g. "I am in contact with everything."), the second factor describes experiences of ego loss and deconstruction (e.g. "My world-view is falling apart."), the third factor includes psychopathological experiences (e.g. "I am controlled by alien forces."), and the fourth factor is pertaining to dreams (e.g. "I dream of future events which occur afterwards."). After six months the retest reliability of the scales ranged between r = .66 (factor III) and r = .89 (factor I).

The analysis of correlations between our questionnaire and other scales, as well as SEM models (Structural Equation Models; AMOS), confirm the hypothesis that spiritual and psychopathological experiences represent different classes of experiences, and that they can be phenomenologically separated by questionnaires. Although the effects of exceptional human experiences on mental health show quite ambivalent outcome effects, a lot of spiritual practices techniques (e.g. meditation, contemplation) seem to adjust the focus on positive outcome effects, while negative impact is apparently suspended.

Address for correspondence: Niko Kohls,
University Hospital Freiburg, Dept. Environmental
Medicine and Hospital Epidemiology,
Hugstetter Str. 55, D-79106 Freiburg, Germany.
E-mail: niko.kohls@uniklinik-freiburg.de

AGAPÈ: GROUP TELEPATHY. A LONG-TERM EXPERIMENTAL SERIES

Bernard M. Auriol[1], Frédérick Garcia, Laetitia Puech[2], Sylvie Lagrange[2], Corinne Morer[2], Myriam Campardon[2], Olivier Rabat[2], Sophie Valentin[2], Olivier Perrin[2] and Eve Leconte[2]

[1]*Institut Métapsychique International, Paris*

[2]*Université Toulouse I and Université Toulouse III*

ABSTRACT

We hypothesized that a majority vote process would favor the ESP success between two groups by enhancement of the signal to noise ratio. We supposed that the repetition of the trials and the use of an individual and group feed-back could improve the telepathy between these two groups of individuals, and obtain results higher than chance expectation.

The participants consisted of 418 unselected volunteers, who participated either in the 'agent' group or the 'percipient' group. The agent group and percipient groups were each placed in acoustically isolated rooms. The size of the agent group varied from 0 to 15 persons and the size of the percipient group from 1 to 16. We also varied several target parameters: target-type (pictures, words), and size of target packs (judging packs consisting of 2, 3 or 5 targets).

For each trial, the computer randomly selects a target from a database of pictures or words, and displays it to the agents. In an isolated room, percipients had to guess the target among randomly selected decoys, using an individual keypad connected to the main computer. Each individual guess, the target, the date and hour of the trial, were recorded on the main computer.

Over the course of seven years, participants contributed a total of 240 sessions, involving 27,845 collective trials, which represent over 250,000 individual trials.

The Overall hit rate was at chance, both for individual trials and for collective (majority vote) trials. Some interesting variance results are noted, though their interpretation remains ambiguous.

The seemingly randomness of our data could belong to the Type II error. This is why we undertook a post hoc covariance analysis on the collective trials which deviated significantly from expectation ($p < .05$). This study made it possible to highlight some interesting parameters, but the variable linked to better results are also linked to worse results. Because that could be linked to a stacking effect, we used a resampling method to control it. The creation of faked targets allowed us to select new significant collective trials. The regression conducted on these resampled data gathered results very similar to those excerpted from the genuine data. We observed the same thing about the sociometric data. Indeed, the replacements of the genuine targets by fake targets does not fundamentally change the results of the covariance analysis: Thus, we are most probably facing a stacking effect.

INTRODUCTION

This paper presents a group telepathy long term experiment. The previous literature offered mitigated results.

Osty (1932) and Barker et al. (1975) reviewing the existing literature noticed that the tests involving groups generally tend to produce "psi-missing", or even purely random results. Haight et al.(1978) attempted to design a method that would be a compromise between an individual approach and a group approach but they were not able to get any significant effect. Milton's meta-analysis (1994) of eight studies (ESP in forced choice), gathering more than one and a half million of individual trials, comes to the same conclusion: The mean effect size was very small, and negative; the overall cumulative outcome of the studies was non-significant. Nevertheless, we should point out that in these experiments participants had been recruited via the mass media and they did not strongly believe in their ability to be good percipients. More important, percipients were not emotionally connected to the agents whereas anecdotal evidences (McConnell, 2004) suggest that affective links are necessary for high Psi performance.

On the other hand several other protocols got positive results: For instance, Barker et al. (1975) and Dalkvist & Westerlund (1998). Similarly, Musso and Granero, (1981) got very significant results and vote success was more impressive than individual rate. Carpenter (1988, 1995) used an ingenious protocol to get Psi information from group interactions. He also got significant results.

Honorton & Ferrari's (1989) meta-analysis of forced-choice precognition studies rejected the null hypothesis ($p< 10^{-24}$). Similarly, several authors have showed correlations between group emotional events and Random Number Generator outputs (see e.g. : Dunne, 1991; Nelson, 1996; Bierman, 1996; Radin et al., 1996; Rowe, 1998 ; Radin, 2002).

We hypothesized that a procedure using repetition of trials and individual and/or group feedback could improve telepathy between two groups because of the redundancy (Kennedy, 2003) associated with a majority vote process. If there are enough individual hits to pull the overall results out of the chance level, it is expected that majority vote technique itself be significant. Yet Brier et Tyminski (1970) paradoxically found that clear cut majorities were less accurate to predict success than slim majorities. Finally Tart (1980) was able to improve results by replacing a single agent by a group.

Some parameters are known to influence the individual psi results: sheep would better succeed than goats (Schmeidler, 1958; Auriol, 2003b), the people practicing an ASC would succeed better than those who do not practice it (See e.g. Tart, 1978, Morris et al., 1993), a spontaneous guess could be more favorable than hesitations, children could better succeed than adults (Randall, 1972) and female better than male (or differently: Freeman, 1970). An instruction given to the agents on the way of sending the target would be favorable (Warcollier, 1926), close links between agents and percipients would be more favorable than an absence of bond (Schmidt et al., 2001, Sheldrake, 2003). Some characteristics of the Lüscher Color Test (Lüscher, 1949), meaning intuition, feeling, extraversion, could be predictive of ESP success (Sargent, 1981; Morris, 1993). Learning could improve the results along the time, or an effect of decline could attenuate them gradually (see e.g. Tart, 1966; Auriol, 1973).

At the historic beginning, Rhine used a constant judging pack of the five Zener symbols. Later, the ganzfeld taught us it is better to use changing potential targets, and images instead of symbols. Otherwise, Barry (1971) used words-targets. In the same way, see Auriol (1973).

AIMS AND PREDICTIONS

We mainly tried to apply several processes of operant conditioning (i.e. individual and/or group-feedback) to a group (agents and percipients) with the prediction of obtaining a response gradually improved and results higher and more stable than those which one can await from a couple of subjects.

Main predictions

The individual hits number would be above the chance expectation. The Resulting of Vote hits would be above the chance expectation. The individual and group feed-back would improve the results. Variance of the number of misses between two consecutive hits would be bigger than the chance expectation. Usage of a variable judging pack could be better than a constant one. Images would be targets better than words.

Secondary predictions

Results would be a function of the number of percipients and of the number of agents, reciprocal knowledge between participants (specifically between agents and percipients) would be favorable, sheep would succeed better than goats, ASC practitioners would succeed better than others, a spontaneous guess could be more favorable than hesitations, children could better succeed than adults (Randall, 1972) and female better (Freeman, 1970 about precognition) than male. Pictures would be easier to transmit than words. An instruction given to the agents on the way of sending the target could be favorable. Some characteristics of the Lüscher Color Test (Lüscher, 1949), meaning intuition, feeling, extraversion, could be predictive of ESP success.

Post Hoc analysis:

We were not able to validate those predictions (except, somewhat, about variance) and we tried to detect some clues about the variables which would favor hitting or missing. We used a multivariate analysis limited to the significant collective trials. But the attractiveness of a "potential target" or spontaneous unplanned ESP phenomena between percipients (Schmeidler and Goldberg, 1973; "mental contagion" of Warcollier, 1926 or "pirate effect": Auriol, 1995), could distort the results (stacking effect). An available parade would be to use the Greville correction (1944). However, this tool, in the context of our study, would be very laborious (Cf. Thouless & Brier, 1970), nearly impossible to deal with. We thus decided to use a resampling method (like Randall, 1972): We replaced the genuine targets by fake targets. This creation of faked targets allowed us to select new significant collective trials. Then, we were able to assess if this manipulation did fundamentally change the results of the covariance analysis (weak stacking effect) or did not (strong stacking effect) (for more details, see Auriol et al., 2004).

METHODS

From December 1993 to January 2001, we carried out 240 telepathic ESP group sessions, made up of 27,845 collective tries (that is more than 250 000 individual tries). The population of participants, who all volunteered and were recruited without special method (friends, acquaintances, ads in papers), consisted of 418 persons. The size of the agent group varied from 0 to 15 people, the size of the percipient group from 1 to 16. We varied a number of parameters, particularly the target's type (pictures, words), and the possible answers' number (2, 3 or 5). The agents and percipients were in two acoustically isolated rooms. Agents had to look at a screen which displayed pseudo-randomly selected targets. Percipients had to guess the target pressing a key on an individual keypad. The computer gave a feedback, or not, to the group (lights) and to the individuals (displaying the winners names or the target). Everything was recorded.

Participants

The population of participants, who all volunteered, were recruited with no special method (friends, acquaintance, media). There were 274 female individuals and 145 male individuals, amounting to 418 persons.

Using questionnaires, we collected complementary information on the participants: age and sex, degree of their adhesion to psi phenomena, their practice of a kind of altered state of consciousness, the results of Max Lüscher color–test (1969). Before sessions, using Likert-like sociometric scales (Likert, 1967; Barnett, 1991), the percipients had to answer on their individual keypad the following questions displayed on their collective screen about each of the other participants (either agent or percipient): "*Do you know this person?*" [mark from «*not at all*» (1) to «*very well*» (5)] and *"are you on the same wavelength as this person?"* [mark from «*not at all*» (1) to «*very well*» (5)] .

Participation was quite irregular from one session to another. Some came nearly at every session, some others, just one or more times. The number of agents varied from 0 to 15 persons, the number of percipients varied from 1 to 16 persons.

Hardware Material

We used one, two or three PCs (in accordance with the needs of our three protocols). They were inter-connected thanks to null-modems and serial RS232 COM ports, using UART.

One of the two Com Ports of the main PC, situated out of reach of the participants in the compartment between the (agents' and percipients') rooms (see Fig. 3), was connected to the network especially designed for the experiment. It was built by Mr JPH Croset , SD Labo (Genève) using the intelligent motherboard Kit Velleman High A K2612 and the extension board for twin-PCB motherboards K2631 High Q (Fig. 2, look at the bottom on the left). It was connected to an output card K2609 and an input card K2611 to connect up to 16 individual keypads, allowing 16 percipients to give their individual guess (Fig. 1).

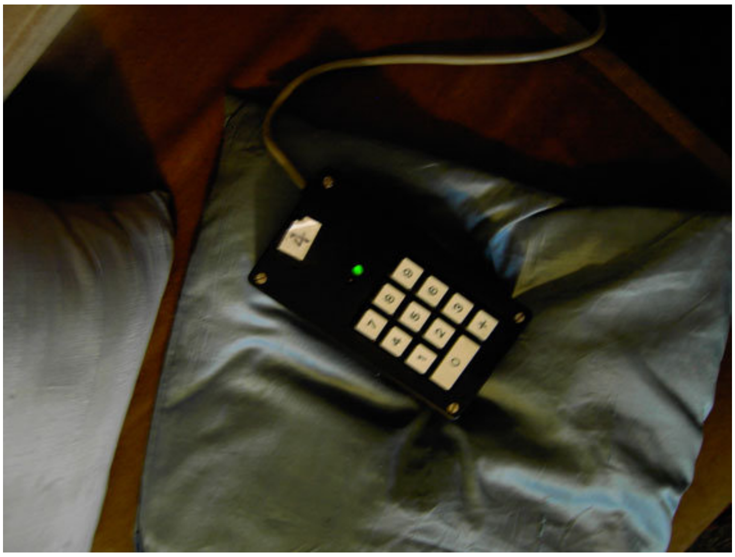


Fig.1: One of the 16 keypads

Fig. 2: The two PCs devoted to display targets in the percipients room

The three feed-back lights were connected to the Velleman System via three electrical relays switching three colored lights (on the right of the screens). The main PC (no photo) was situated out of reach of the participants. Its screen, as the traffic light giving feedback to the agents, was situated in the Agents' room for displaying the target to them.

The programs were written with Turbo-PASCAL 7, except some batch DOS files.

Map of the experiments site

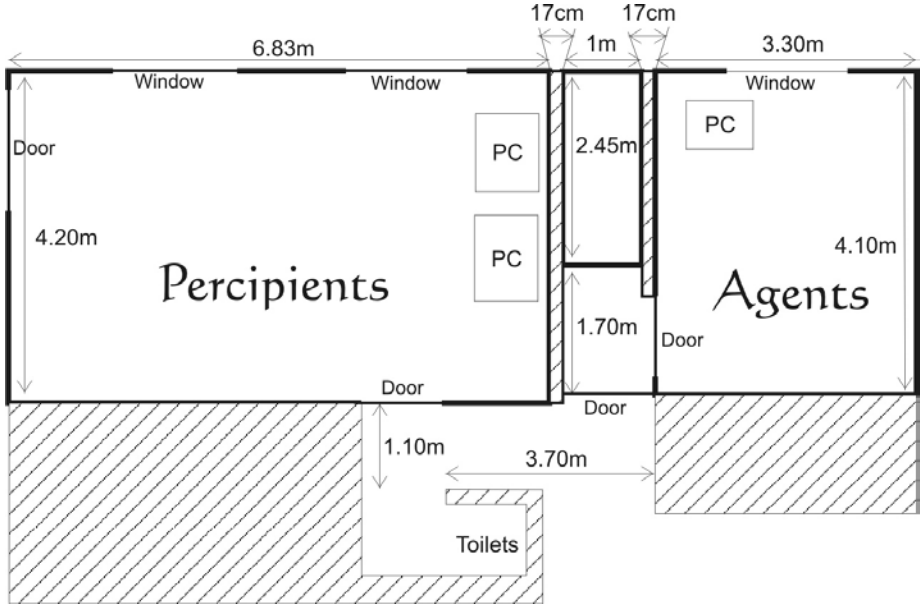


Fig. 3: Map of the experiment site N°1 (Blanchard Street)

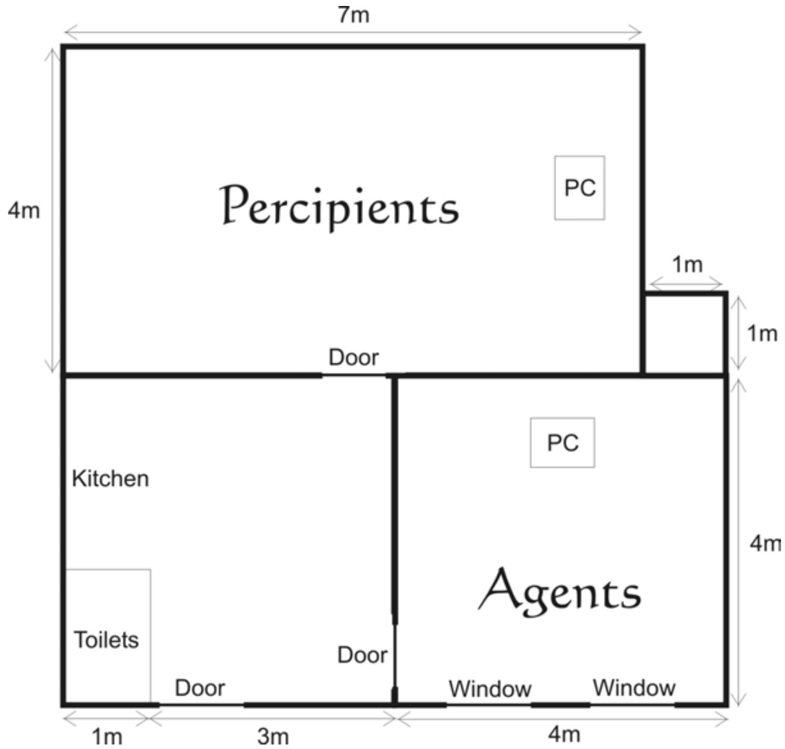


Fig. 4: Map of the experiment site N°2 (Riquet Street)

The sessions took place in two different sites. Most of the sessions (N° 2-16 and 97-240) were at the professional office of Dr. Bernard Auriol (Fig. 3: Blanchard Street); for 80 other sessions (from N° 17 to N° 96) we had the opportunity of using a more beautiful "laboratory" (Fig.4 : Riquet Street). In both cases phonic isolation was quite good: Thick walls for Riquet Street, double partition walls coated with "calibel" (see http://espace.isover.fr/crm/ac_sonic.htm); "doors" refers to 4cm thick solid wood doors with no recessed panel.

Target Materials

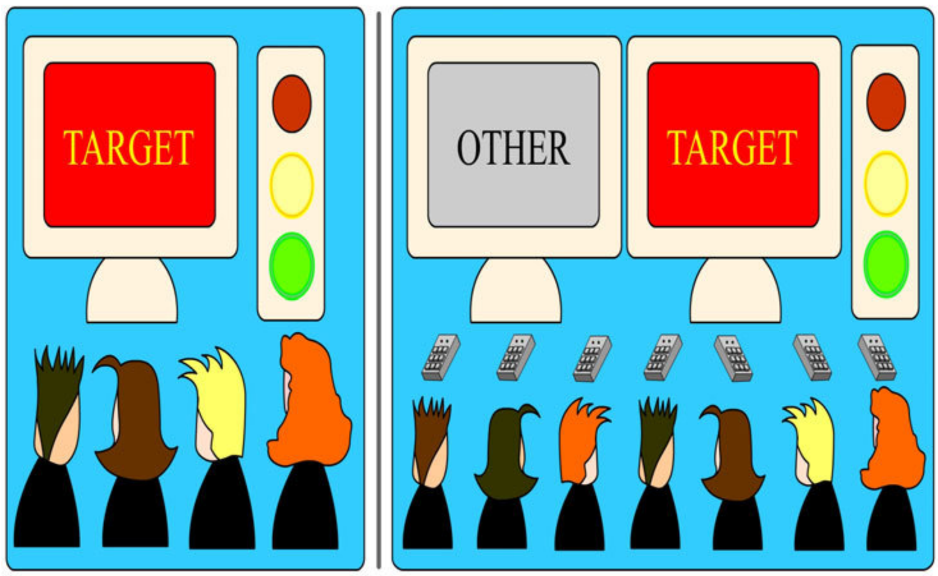
Overall three main protocols were implemented. It seemed interesting to compare the condition in which the potential targets are always the same to the condition in which they change for each trial. Similarly, are images better potential targets than words? To test these questionable differences, variations of layout were used for practical reasons depending on the nature of the available material and the layout of the premises.

When we have wanted to test if a change of the judging pack, for each call, would be favorable, we had to switch from five potential targets to three. Indeed, when the five targets were always the same ones along the whole session, we were able to post them at the beginning on a paper-board; on the other side, when the potential targets had to vary with each test, it became useful to post them on a monitor. To obtain a clear posting, which would be perceptible from the most remote place of the percipients room, we had to reduce the number of possible targets to three.

In the same way, when we wanted to apply the same type of protocol to target-images (instead of words), we had to use, for the same reasons of visibility, one screen for each potential target. So that, to avoid too many hardware complications and to reduce the cost, we had to limit the number of potential targets to two (two screens).

That is why, three different target-types and corresponding judging procedures were used over the sessions:

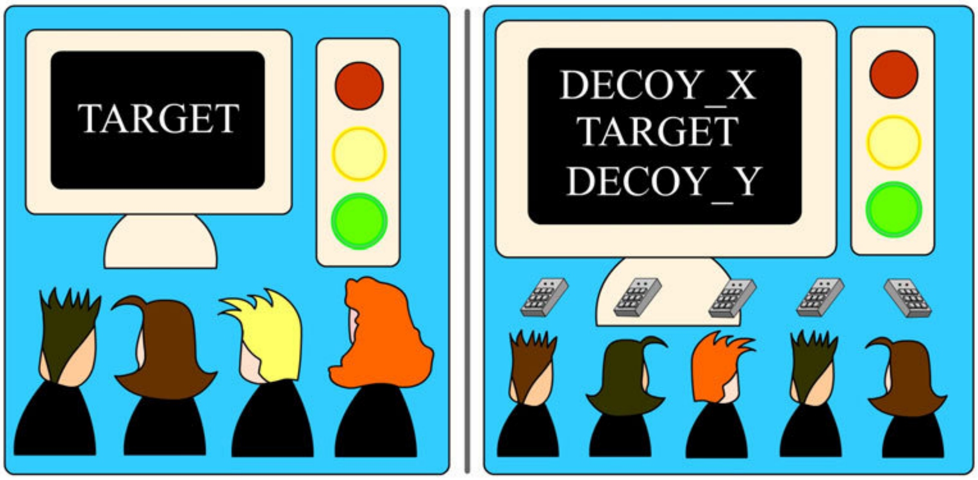
A. Two Pictures (fig. 5) : For each trial, the computer randomly selects, thanks to the Pascal function "Random", a target from among 1,500 pictures ("1500 Photos", 1996), and displays this to the agents. The percipients must choose between two pictures (the target, plus another randomly selected picture), displayed on their screen.



Agents in front of their screen	Percipients (the target is displayed at random on the left or on the right)

Fig. 5: Experimental design ("two pictures" protocol): two acoustically insulated rooms

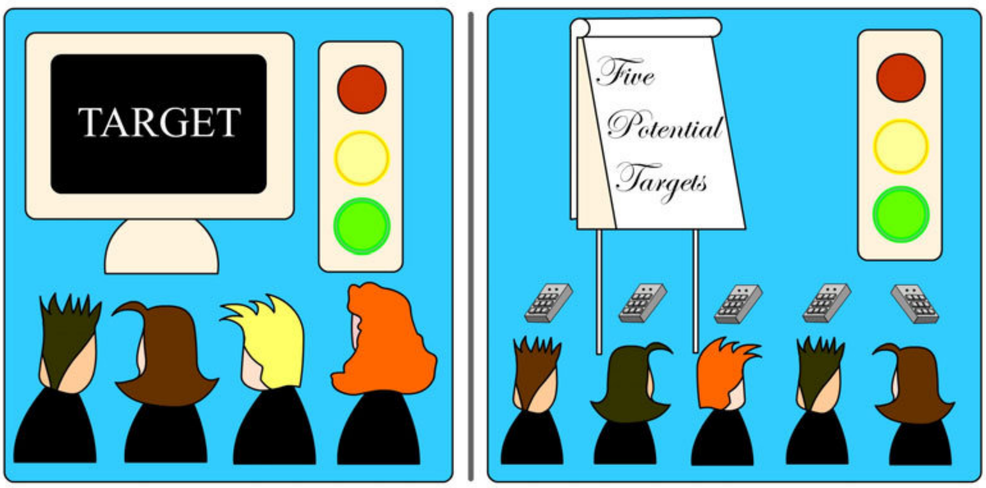
B. Three words (fig. 6): We used a file of 1,000 very common words (extracted randomly from "Le plus petit Larousse", 1946).For each trial, the program picks a list of ten words among a file of 100 lists (1,000 words). Then it selects randomly one of these ten words plus the two next words of this list. One of these three words is randomly picked (Turbo-Pascal function "Random"): it will be the target, the two other words, the decoys. These three words are displayed to the percipients. Everyone gives his/her answer on an individual keyboard.



Agents	Percipients (the target is displayed at random on one of the three rows)

Fig. 6: Experimental design ("three words" protocol): two acoustically insulated rooms

C. Five words (fig. 7): at the beginning of the session, all the participants (agents and percipients together) brainstorm a list of five words which will be used repeatedly during the whole session. For each trial, one of these five words is randomly selected thanks to the Turbo-Pascal function "Random" and displayed to the agents. The percipients give their answer on their individual keyboard.



Agents	Percipients' room: the potential targets, chosen by the whole group before the action, are displayed on a paperboard

Fig. 7: Experimental design ("five words" protocol): two acoustically insulated rooms

Procedure

At first, after a brief explanation about the experiment and its protocol, the participants had to complete a questionnaire: surname, name, age, sex, phone number, choice of participating as agent or as percipient and also label of the chosen keypad (Fig. 1 => "4"), belief in ESP and practice of ASC.

Here is the sheep-goat question we used:

"Are you convinced of the existence of telepathy, precognition, clairvoyance, and psychokinesis? (for each, three possible answers: No, I do not know, Yes)".

Here is the ASC question:

Do you practice a relaxation, meditation or orison method? (Yes or No for each).

Then the experimenter filled in a computerized questionnaire which records Surname, Name and Keypad Label (Fig. 1) if appropriate.

Then the percipients sat comfortably nearby their proper keypad, and had to answer on it at the sociometric question:

"Do you know X Y (name and surname)?,". At each occurrence of this display, the named participant stood up to introduce oneself. The percipients gave a mark from "not at all" (1) to "very well" (5) which was automatically recorded.

The "agent" and "percipient" groups were located in two acoustically and optically isolated rooms (Fig. 1 & 2).

The participants chose their role (agent or percipient) at the beginning of the session, and kept this role throughout the session. The number of participants in the agent-group varied from 0 (no agent) to 15 persons, while the number of participants in the percipient-group varied from 1 to 16 persons. [Once, there was no agent: This was requested by one of the participants believing that telepathy is the same process as clairvoyance].

At the beginning of each trial, the target and decoys were selected pseudo-randomly thanks to the procedure "randomize" followed by the "random( )" function of Turbo-Pascal v.7 (Stermole, 1998).

For the picture-trials, the target was selected from a 1,500 images CD-Rom; for the three-words trials it was extracted from 1,000 very common words and for the five-words trials the target was selected from the five words chosen at the beginning of the session by all the participants (agents with percipients).

The target (word or picture) was then displayed on the agents' screen until the trial was over.

In terms of instructional set, in some cases agents were free to "send" the message in whatever way they judged best, in other cases they were given instructions as to how to send it. In the latter cases, they were either instructed to focus on the target itself, or else to focus on one (or several) percipients.

Meanwhile, after a brief latency (blank screen), each member of the percipient group observed, on the collective screen, the judging pack (consisting of 2, 3 or 5 items; see above), and used his/her individual keypad (Fig. 1) to input his or her guess. If the majority had chosen the right answer, it was counted as a collective hit, otherwise a miss.

After the judging was closed a feedback was given.

- Individual feedback: each receiver is informed of his/her personal hit or miss, by displaying on the screen either his/her name or the target.

- Feedback for the group: when the trial was over, if the vote was a hit, the lights were swtched on (traffic lights' style: see on the right of the displays, fig 5, 6 and 7) in the senders' room and the agents'. For some sessions, one light was switched on even if the majority gave a miss provided that the number of right individual answers be above the mean chance expectation. The number of lights on depended on the strength of the majority.

At the end of the session, the participants could see on the screen the individual and group results, with stars for significance; they were asked to give their feelings (Auriol, 2001a).

In order to determinate if there are conditions favoring ESP results, we varied some other details in the protocol (Auriol, 2001b) : The « allowed maximum time for the vote » ranged from 20 to 120 seconds. This period may be cut short: The time left to vote may be shortened after half of the percipients had voted.

A total of 75 trials (or more), as described above, are collected for a single session. The target materials are always the same, e.g. images, during a session. All session data (agents' names, percipients' names, rank assigned to the trial, source-set, decoy subset, target, date, display time and duration, time taken to answer for each receiver) are saved on disk. From December 1993 to January 2001, we carried out 240 sessions of group telepathy, involving 27,845 collective trials; these were based upon over than 250,000 individual trials. It should be noted that, as the receivers were not obliged to make a selection (they could "pass") a lot of missing values were generated.

RESULTS

Percentage of individual hits

If the choice of a percipient matches the target, it's an individual hit. For the protocol with two pictures (respectively three words and five words), we checked if the percentage « $\hat{p}$ » of right answers was equal to p_o = 1/2 (respectively 1/3, and 1/5), that is compatible with the null hypothesis.

The hypotheses in this test are: $H_0 : p = p_o$ versus $H_1 : p \neq p_o$

The statistic used is : $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ where n is the number of trials associated to each protocol. It

asymptotically follows a normal standardized distribution (table 1).

	Number of individual trials	Percentage of hits	P-value
2 pictures	27,081	49.94 %	0.83
3 words	102,634	33.34 %	0.95
5 words	120,347	20.13 %	0.27

Table 1: Hit rate of individuals

Influence of individual characteristics on individual hits

A hit mark can be given to each percipient for each session he took part in. This mark is based upon the percentage of hits he got for each session. In order to include in the analysis the individual percentages got during the sessions with two pictures, three or five words, we apply the following transformation (Saporta, 1978):

$$h = 2(\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p})$$

We get ***h*** for each person. We try to evaluate whether ***h*** takes significantly different values according to recorded individual characteristics as we had hypothesized being of interest (see introduction): year of birth, sex, Lüscher's test, degree of adhesion to psi phenomena (telepathy, clairvoyance, precognition or psychokinesis) and practice of relaxation, orison or meditation.

Then we compare the means of the random variable ***h*** distribution between different groups of individuals, discriminated according to the following criteria:

Age: the participants were classified in four quartiles, according to their date of birth: the persons born before 1935, those born between 1935 and 1943, those born between 1943 and 1960, and those born after 1960. Age could be measured as a correlation, but due to the recruitment it seemed to us better to use quartiles.

Color: groups were determined by the presence or not of each of the eight colors of the Lüscher color test (blue, green, red, yellow, purple, brown, grey and black) among the four favorite colors of the participant.

Sheep/Goat: groups defined by their degree of adhesion to psi phenomena (telepathy, clairvoyance, precognition or psychokinesis): strong adhesion /moderate adhesion / no adhesion of the participant.

ASC: group of individuals practicing relaxation, prayers or meditation, versus the group of individuals using none of these.

The variance analysis (and the Fischer test associated to it) allows us to compare the means of these different groups: none of the different individual variables did significantly differ from the chance expectation: The P-values range from 0.16 to 0.81 (See APPENDIX, Table 2 and 3). We cannot associate these characteristics to a wider success of the participants.

If we repeat the analysis taking this time into account only the persons who got a significant hit rate, we reach the same conclusion. However, if we take into account only the persons who got a significant miss rate, some characteristics seem to have an influence on their mark (Table 3).

Age	Fischer's Test Statistics	P-value
Date of birth	6.06	0.002

Sheep / goat questionnaire	Fischer's Test Statistics	P-value
Clairvoyance	5.42	0.02
Psychokinesis	6.60	0.003

Table 3: significant individual characteristics regarding the “psi-missers” sub-population

Believing strongly in psychokinesis or clairvoyance tends to increase the psi-missing compared to moderately believing in these phenomena; the youngest participants got the highest psi-missing.

Percentage of collective hits

For each collective trial, if the majority of percipients choose the target, it’s a collective hit (group hit).

Number of Potential Targets →	**Two**	**Three**	**Five**
Expected mean	0.500	0.333	0.200
Observed mean	0.498	0.329	0.202

Table 4: hit rate of collective trials (vote results).

Variance

We could suppose that high percentages (those higher than expected by chance, or Psi-Hitting), and low percentages (Psi-Missing) compensated one another. Under this hypothesis that there are fluctuations between attitudes in Psi-Hitting and Psi-Missing, it can be interesting to test the variance of success. In order to evaluate in the most precise way the variation of success in relation to time, we can note the number of misses between two consecutive hits (interval), and check if the variance of these intervals is random or not.

Our use of the vote makes the study of the individual success variance more difficult (Hyman, 1992). We will therefore limit our study to the variance of collective hits (Thouless & Brier, R.M., 1970).

Nb of potential targets	Nb of hits (N intervals +1)	Observed variance	Expected variance
two	1463	**1.93***	*2*
three	3512	**6.42***	*6*
five	777	20.05	20

Table 5: variance of intervals between consecutive successes (got by vote)

When the percipients have to choose between two potential targets, the results tend to gather around the mean. The variance of intervals is significantly lower than chance expectation ($p<.05$ table 5).

Conversely, for the protocol with three potential targets, they tend to depart from the mean. The variance is significantly higher than chance expectation ($p<.05$ table 5).

However, in the case where the percipients have to choose between five potential targets belonging to a same reiterated set, the variance isn't different from the expected variance (n.s. table 5).

Strength of the vote and success

We tried to learn if a better agreement among the voters on the answer to give leads to more success. We call "strength of the vote" for a given trial, the ratio of individuals who agreed upon an answer (majority vote), whether the result is good or not.

The null hypothesis is: "the percentage of hits amounts to chance expectation, whatever the strength of the vote". We can suppose that, if some answers are not due to chance but to ESP, this should have an impact on the majority: Strong majorities could be more linked to success than weak ones. In fact the calculations carried out on the sessions made us reject this hypothesis: strong majorities didn't get better results than weak ones (See APPENDIX, Table 6, Fig. 8, Table 7, Fig.9 and Table 8, Fig. 10) .

Post Hoc investigation: Covariance analysis

The seemingly randomness of our data could belong to the second kind error. This is why, on a purely heuristic basis, and under the assumption that ESP would be (within the framework of our protocol) a scarce phenomenon with a weak impact and/or unspecified sign (sometimes towards Psi-Hitting, sometimes towards Psi-Missing) we undertook a covariance analysis on the collective trials which deviated significantly from expectation ($p < .05$). We are well aware that we can expect 5% of the collective trials to be different from chance under the threshold $p=0.05$, but we don't foresee that a regression on these trials leads to identify the parameters which would increase or decrease the difference from chance (H_o).

We are working at the level of a collective trial. For each collective trial, the proportion of individual hits is used as dependent variable. For the three protocols, this proportion has the following distributions.

- **Two pictures:**

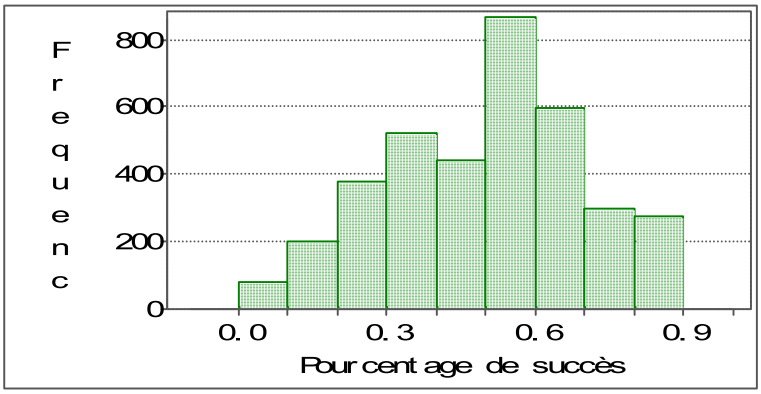


Fig. 11: Frequencies of individual hit rates into the group trials ("two-pictures protocol")

- **Three words:**

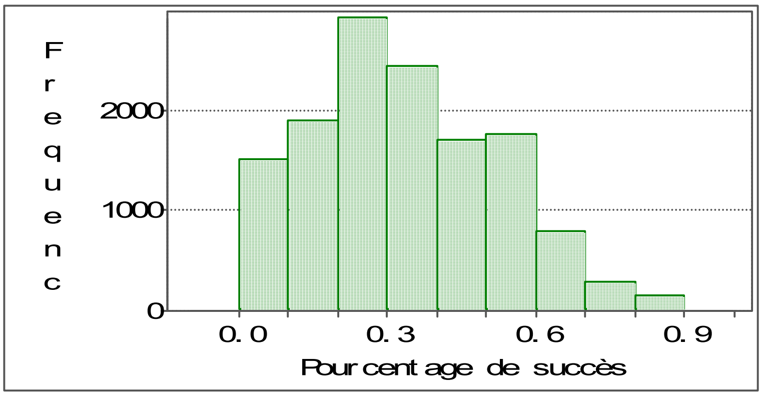


Fig. 12: Frequencies of individual hit rates into the group trials ("three-words protocol")

- **Five words :**

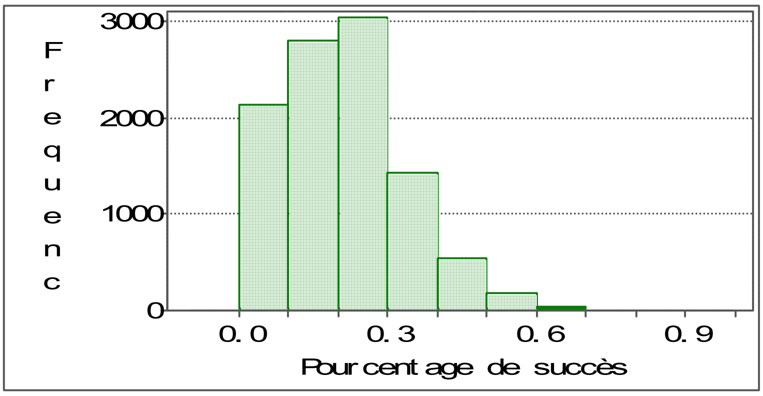


Fig. 13: Frequencies of individual hit rates into the group trials ("five-words protocol")

The distribution of hits rate is centered on 1/2 for two pictures (Fig. 11), on 1/3 for three words (Fig. 12), or on 1/5 for five words (Fig. 13), that is on the percentage of right answers under the null hypothesis.

In order to test the effect of each variables' modality, we used a transformation of the « percentage of hits » to be able to compare the results for the protocols with two pictures, three or five words (Cohen, 1977, 1988; Saporta, 1978).

$$h = 2(\arcsin\sqrt{\hat{p}} - \arcsin\sqrt{p})$$ where:

- $\hat{p}$ is the percentage of right answers in the trial
- p is the expected percentage (0.50 for two pictures, 0.33 for three words and 0.20 for five words).

If we look into the quality of the trials, we can test if the percentage of hits received for each trial is significantly lower than chance, significantly higher than chance, or equal to chance thanks to a test of Khi². The statistic of this test, calculated for each trial, is

$$\left(\frac{\hat{p}-p_0}{\sqrt{(p_0(1-p_0)/n)}}\right)^2 \sim \mathrm{Khi}^2 \text{ (one df)}$$

From the twenty-five variables we considered, eight were selected by a stepwise corrected analysis (SAS software). In every case, the Fisher test was significant with a p value close to 0.0001.

	Variable	Estimate	
	Collective trials	**Higher than chance**	**Lower than chance**
	Intercept	+1.06	-1.12
Qualitative Variables	The transmitters were given an instruction	+0.10	...
	The participants chose the targets' list	-0.34	+0.26
	There was some kind of group reward	-0.13	...
	The time left once half the receivers answered = the time already spent	+0.27	-0.02
	The time left once half the receivers answered = twice the time already spent	+0.30	-0.12
Quantitative Variables	Time to answer	-0.01	+0.01
	Ratio {nb of transmitters / nb of participants}	-1.46	...
	Ratio {nb of transmitters / nb of receivers}	+0.96	...

Table 9: results of covariance analysis on the significant trials at $p<0.05$ (after Auriol, 2003c)

This study (Table. 9) made it possible to highlight some interesting parameters : hitting and missing seem to be enhanced when the agents were told to concentrate either on some specific percipients or on the target; hitting and missing seem to be enhanced when the time to answer is shortened.

We studied separately the effect of the sociometric bonds on the results (Table. 10). As we noticed, the receivers gave each participant a mark going from 0 to 5, answering the question: "Do you know this person? (Likert scale from "not at all" to "very well")".

We made two fractions:

$$RecGr = \frac{\text{Receivers' Mark}}{\text{Group's Mark}} \qquad TransGr = \frac{\text{Transmitters' Mark}}{\text{Group's Mark}}$$

where “Receivers’ mark” is the mean of the marks given by the percipients to the other percipients, and “Transmitters’ Mark” is the mean of the marks given by the percipients to the agents. The « Group’s Mark » is the mean of all the marks.

For the significant trials (threshold 5%) with higher results than expected (Psi-Hitting), the table had 641 observations. The adjusted R^2 equals 10.8% (« small » effect according to Cohen’s convention) (Table 10).

Variable	**Estimate**	**Success**	**Distance from chance**	**P-value**
RecGr	-1,54	-	-	<0.0001
TransGr	+0,58	+	+	0.0003

Table 10: REG procedure on sociometric bonds (after Auriol, 2003c)

For the significant trials (threshold 5%) with lower results than expected (Psi-Missing), the table had 355 observations and the estimate corresponding to RecGr was 1.29. TransGr was not significant. The sociometric bonds among percipients would reduce the distance to chance expectation whereas the sociometric bonds between agents and percipients would increase this distance.

What about the Stacking effect?

In every case, the variables linked to better results are also linked to worse results. In that way, we could suspect that the significance is increasing only because the receivers give more similar non-independant answers, independently of the transmission process. For each collective trial, each member of the percipient group chooses from a common set of potential targets. If one potential target has attractive or repulsive characteristics, there will be a surplus of votes in favor of this potential target or away from it, regardless of any telepathic phenomenon. If it happens that the real target matches an “attractive potential target”, we’ll have a fake “hit”, conversely, if the real target matches a “repulsive potential target”, we’ll get a fake “miss”. Indeed this result could be linked to a “stacking effect” (Caroline Watt, 2003).

An available parade would be to use the Greville correction (1944). However, this tool, in the context of our study, is very laborious (Cf. Thouless & Brier, 1970) and is nearly impossible to deal with. We thus decided to use a resampling method like did Randall (1972).

Our resampling study of the stacking effect (for more details, see Auriol et al., 2004) shows that replacements of the genuine targets by fake targets does not fundamentally change the results of the covariance analysis. Indeed, the creation of faked targets allowed us to select new significant collective trials. Regression conducted on these significant collective trials gathered results very similar to those excerpted from the genuine data. We observed the same thing about the sociometric data. (see APPENDIX: Tables 11, 12 and 13) Thus, we are most probably facing a stacking effect.

DISCUSSION

Concerning the frequency of individual and collective successes, our results are consistent with the chance assumption. We were not able to obtain any improvement of the signal to noise ratio based on the redundancy provided by the majority vote. Strong majorities did not give better results than weak ones.

Our collective experiments yielded lower results than these reported by other researchers, in particular on protocols with pairs. There might be several reasons for this negative result: Our participants were recruited without any selection, our too rigid protocol would obstructed the manifestation of Psi, or the group situation and some micro-sociological phenomena would inhibit the manifestation of ESP phenomena (for unspecified reasons).

One could invoke a strong impact of a negative Psi-experimenter effect. But Bernard Auriol (1974, 1994, 2003a), the main experimenter in "Agapè", has conducted other Psi works with quite good results for some of them.

Another potential explanation could be related to a topographic proximity question. A generally accepted idea, documented rather well, is that ESP hitting did not depend on distance (Sheldrake & Smart, 2003; Braude, 1979; Stevenson, 1970). Nevertheless, Osis (1965) found a decrease in effect size over distance when examining a group of 22 five-alternative forced-choice studies. Direct empirical tests of the effect of distance generally show a decline in effect sizes with distance (Milton, 1994). So topographic distance between agents and percipients could be weakening the effect size. The design of our experiment brings percipients closer one another, and agents as well between themselves. Yet our design moves the percipients away from the agents. We could suspect that the effect of distance be of interest mainly for very short ranges and vanishes for larger ranges. If this distance effect was proportional to a rather high power of its length, the transmission would be very important between percipients and dramatically weakened between agents and percipients. Even though surprising, this hypothesis is consistent with other findings regarding time: the effect size of precognition would be decreasing over the time (Honorton and Ferrari, 1989). Thus we envisage that effect size would depend upon space-time distance between agent and percipient.

Moreover, as the percipients are put together in only one room, they interact in several ways: by smell, sight, sounds and why not, ESP.

The variance evaluated with the method of the intervals between two consecutive successes of the vote was significantly different from chance for two of the three protocols; but the interpretation of this result is not obvious: One possibility is that there could be individual or group psychological fluctuations. These fluctuations, following an unspecified law, would alternate phases of psi-hitting and psi-missing; the overall performance would then be random, since successes and failures would compensate one another. This could be an unconscious management of approval and refusal. This management would reduce any telepathic manifestation. That would explain - as suggested by Kennedy (2003) – the phenomenon of decline and the "actively evasive, unsustainable" character of Psi.

Our study of individual characteristics did not yield more significant correlations with hits rate, but we have to explore some new questions: It would be interesting to look at personality characteristics of those choosing to be sender and those choosing to be receivers. We also plan to do a *post hoc* analysis on the decoys that the computer picked out in each session. Re-analysing our dataset using scoring extremity as a variable might yield some interesting results.

We carried out a multivariable analysis, regarding variables which we had supposed suitable to strengthen or weaken the Psi at work. An resampling method for tracking the stacking effect allowed us to discard these post hoc conclusions (Puech et al., 2003; Auriol, 2004).

CONCLUSION

The results didn't fulfill our expectations, especially regarding a possible improvement of the ratio signal to noise linked to the redundancy got from majority vote. This way of carrying out the experiment, not only didn't strongly increase the Psi-Hitting as we expected, but seems to have made all the results random, for either individual answers or answers obtained by vote. We especially noticed that, unlike what we expected, strong majorities didn't get better results than weak ones.

The variance evaluated with the method of intervals between consecutive hits of the vote significantly departed from chance expectation for two of the three protocols; but this result is very hard to read: The variance is normal, significantly low, or significantly high compared to the expected one, depending on the protocol used.

The multivariate analysis was polluted by a stacking effect and the exhibited variables seem to explain more the strength of the majorities than any link between these majorities and the actual targets.

ACKNOWLEDGMENTS

Grant: Fondation Odier de Psycho-Physique.

Pictures and map: Benjamin Auriol.

English translation: Alexia Fournier; Reviewed by Karen Hansen, Dr Mario Varvoglis and Pr Emmanuelle Auriol.

REFERENCES

I500 Photos, (1996) *SoftKey, CD-Rom Disc*, CDFS, 421 Mo, Genevilliers-Paris 25 Janvier 1996, (c)1992-1995 SoftKey Multimedia Inc.

Auriol, B.M. (1973) Apprentissage et télépathie ? [On-line] Available: *http://auriol.free.fr/parapsychologie/apprenti.htm*

Auriol, B.M. (1974). Parapsychology at school, Montalembert, Toulouse. [On-line] Available: *http://auriol.free.fr/parapsychologie/MontaENGL.htm*

Auriol, B.M. (1992). Proposition à la Fondation de Psycho-Physique, Marcel et Monique Odier. [On-line] Available: *http://auriol.free.fr/parapsychologie/Agape/agapropo.htm*

Auriol, B.M. (1994). Sur l'apprentissage de la transmission parapsychologique entre êtres humains, Revue Française de Psychotronique 4, 3, 211-217, 1994. [On-line] Available: *http://auriol.free.fr/parapsychologie/apprenti.htm)*

Auriol, B.M. (1995). Nouvel élan pour le Projet AGAPE ? [On-line] Available: *http://auriol.free.fr/parapsychologie/Agape/histagap1.htm*

Auriol, B.M. (2001a), Remarques des participants, (relevées à la fin de chaque session Agape) [On-line] Available *http://auriol.free.fr/parapsychologie/Agape/remarques.htm*

Auriol, B.M. (2001b), Variations de Protocole pour le Projet Agape [On-line] Available: *http://auriol.free.fr/parapsychologie/Agape/protocoles.htm*

Auriol. B.M. (2003a). Tests menés sur la précognition à partir d'une observation fortuite. [On-line] Available : *http://auriol.free.fr/parapsychologie/preco.htm*

Auriol. B.M. (2003b). Chèvres et moutons (sceptiques et croyants en l'existence du Psi) [On-line] Available : *http://auriol.free.fr/parapsychologie/chevre/*

Auriol, B.M. (2003c) Agape Experiment: Further Statistical Studies (In Progress), *EuroPA meeting, 14th-17th November 2003, Port-Royal, France.* [On-line] Available : *http://cabinet.auriol.free.fr/parapsychologie/agape/further-stat-study.htm*

Auriol, B.M., Garcia, F. and Puech L. (2004) Comparative Study with the purpose to check the impact of a possible stacking-effect on the Agapè Data. [On-line] Available : *http://cabinet.auriol.free.fr/parapsychologie/agape/stack_covar.htm*

Barker, P.L. Messer E. and Drucker S.A. (1975). Internally-Deployed Attention States : Relaxation A Group Majority Vote Procedure with Percipient Optimization, *Research in Parapsychology, Abstracts and Papers from the Sixteenth Annual Convention of the Parapsychological Association, 1975, 165-167*. [On-line] Available : *http://auriol.free.fr/parapsychologie/redondance/barker75.htm*

Barnett, V. (1991) *Sample Survey principles and methods*. Hodder publisher.

Barry, J. (1971) Journal d'un parapsychologue, Paris, Publications Premières.

Bierman, D.J. (1996). Exploring correlations between local emotional and global emotional events and the behaviour of a Random Number Generator. *Journal of Scientific Exploration,10, 3, 363-373.*

Braud, S. (1979). *ESP and psychokinesis: A philosophical examination.* Philadelphia, Temple University Press.

Brier, R.M. & Tyminski, W.V. (1970). Psi application:Part II, The majority vote technique – analyses and observations, *Journal of Parapsychology, 34, 1.*

Carpenter, J.C. (1988). Quasi-therapeutic group process and ESP, *Journal of Parapsychology, 52, 279-304.*

Carpenter, J.C. (1995). Small Group Interaction as a Medium for ESP Response, *Thirty Eighth Annual Convention of the Parapsychological Association, 80-87* [On-line] Available : *http://auriol.free.fr/parapsychologie/redondance/carpenter95.htm*

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*, New York, Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*, Hillsdale, NJ; Erlbaum.

Dalkvist, J. and Westerlund, J. (1998). Experimental Studies on telepathic Group Communication of Emotions, *Journal of Scientific Exploration, 12, 4, 583-603*.[On-line] Available : *http://auriol.free.fr/parapsychologie/Dalkvist98.htm*

Dunne, B.J. (1991) *Co-operator experiments with an REG device.* Technical note PEAR 91005, Princeton Engineering Anomalies Rsearch, Princeton University.

Freeman, J.A. (1970). Sex differences in ESP response as shown by the Freeman Picture-Figure Test. *Journal of Parapsychology, 1, 37-46.*

Haight, J. Weiner, D. and Morrison, M. (1978). Group testing for ESP : a novel approach to the combined use of individual and shared targets, *Research in parapsychology, Abstracts and Papers from the twenty-first Annual Convention of the Parapsychological Association*, 96-98. [On-line] French translation Available : *http://auriol.free.fr/parapsychologie/redondance/haight78.htm*

Honorton, C. & Ferrari, D.C. (1989). Meta-analysis of forced-choice precognition experiments, *Journal of Parapsychology, 53, 281-308.*

Kennedy, J.E. (1979). Redundancy in Psi Information, Implications for the Goal-Oriented Hypothesis and for the Application of Psi, *Journal of Parapsychology. 43*.[On-line] Available: *http://auriol.free.fr/parapsychologie/redondance/red1.htm.*

Kennedy, J.E. (2003). The capricious, actively evasive, unsustainable nature of psi : a summary and hypotheses, *Journal of Parapsychology, 67, 53-74.*

Larousse (1946). *Le plus petit Larousse*, Paris, Larousse. (i.e. the smallest French dictionary).

Likert, R. (1967). *The human organization: Its management and value*. NY: McGraw-Hill.

Lüscher, M. (1949-1969). *The Lüscher Color Test, translated and edited by Ian Scott. Based on the Original German Test,* Random House, New York (see : *Der Lüscher-Test, Persönlichkeitsbeurtailung durch Farbwahl*, with bibliography of 131 items, 1971, Rowohlt; *french trans,* Aubanel, Avignon, 1973).

McConnell, R.A.(2004) Psi bonds: What are they? Lust is not a sin. [On-line] Available: *http://www.ramcconnell.com/psibonds.htm*

Milton, J. (1994). Mass ESP : a meta-analysis of mass-media recruitment ESP studies, *the Thirty Seventh Annual Convention of the Parapsychological Association, 284-292*[On-line] Available: *http://auriol.free.fr/parapsychologie/redondance/milton94.htm*

Morris, R.L., Cunningham, S., McAlpine, S. and Taylor, R. (1993). Towards replication and extension of autoganzfeld results, *Proceedings of the 36th Parapsychological Association convention, Toronto, Canada.*

MussoJ.R. & Granero M. (1981). Group Gesp experiments tending to yield repeated positive results, Research in Parapsychology, *Abstracts and papers from the twenty-fourth annual Convention of the Parapsychological Association",* *The Scarecrow Press, Inc., Metuchen, NJ & London, pp. 100-103* [On-line] Available: *http://auriol.free.fr/parapsychologie/redondance/musso81.htm*

Nelson, R.D., Bradish, G.J., Dobyns, Y.H., Dunne, B.J., and Jahn, R.G. (1996). FieldREG anomalies in group situations, *Journal of Scientific Exploration,10, 1, 111-141*

Osis, K. (1965). ESP over distance: A survey of experiments published in English. *Journal of the American Society for Psychical Research, 59, 22-42*

Osty, E., (1932). Télépathie Spontanée et Transmission de Pensée expérimentale, *Revue Métapsychique, IMI-Paris, 4, pp.233-256* [On-line] Available : *http://auriol.free.fr/parapsychologie/redondance/osty32.htm*

Puech, L., Garcia, F. and Auriol B. (2003) Etude comparative avec resampling en vue de détecter un éventuel «*stacking effect*» dans la régression sur les données d'Agape. [On-line] Available: *http://cabinet.auriol.free.fr/parapsychologie/agape/2-etude_stacking_effect1-resampling.htm*

Radin, D.I., Rebman, J.M., and Cross, M.P. (1996) Anomalous organization of Random events by group consciousness: Two exploratory experiments, *Journal of Scientific Exploration, 10, 1, 143-168.*

Radin, D.I. (2002) Exploring Relationships Between Random Physical Events and Mass Human Attention:

Asking for Whom the Bell Tolls, *Journal of Scientific Exploration, Vol. 16, No. 4, pp. 533–547, 2002*

Randall, J. (1972). Group ESP experiments with schoolboys, *Journal of Parapsychology, 36, 2.*

Row, W.D. (1998). Physical measurement of episodes of focused group energy, *Journal of Scientific Exploration,12, 4, 569-581.*

Saporta, G. (1978). *Théories et méthodes de la statistique*, Technip, Paris.

Sargent, C.L., Extraversion and performance in extrasensory perception tasks, *Personality and individual differences, 2, 137-143.*

Schmeidler, G. & Goldberg, J. (1973) Evidence for selective telepathy in group psychometry, *Research in Parapsychology, Abstracts and papers from the sixteenth annual Convention of the Parapsychological Association", The Scarecrow Press, Inc., Metuchen, NJ & London, 103-106.*

Schmidt, S., Tippenhauer, G. & Walach, H. (2001) Connectedness Between Participants in an Experiment on Distant Intention. *Proceedings of Presented Papers: The Parapsychological Association Forty-forth Annual Convention, 285-300.*

Sheldrake, R.& Smart, P.(2003). Videotaped experiments on telephone telepathy, *Journal of Parapsychology, 67, 147-166.*

Stevenson, I. (1970). *Telepathic impressions. Charlottesville:* University Press of Virginia.

Tart, C.T. (1966). Card Guessing Tests: Learning Paradigm or Extinction Paradigm? *Journal of the American Society for Psychical Research, 60, 1, 46-55.* [On-line] Available : *http://www.paradigm-sys.com/display/ctt_articles2.cfm?ID=36*

Tart, C.T., Chambers, P., and Creel M. (1980). The neglected sender: preliminary indication that multiple senders may enhance psi performance, *Research in Parapsychology, Abstracts and papers from the twenty-third annual Convention of the Parapsychological Association", The Scarecrow Press, Inc., Metuchen, NJ & London, pp. 107-109*

Tart , C.T. (1978) *Psi Functioning and Altered States of Consciousness: A Perspective,* In B. Shapin & L. Coly (Eds.), Psi and States of Awareness. New York: Parapsychology Foundation, 180-210. [On-line] Available : *http://www.paradigm-sys.com/cttart/sci-docs/ctt78-pfaas.html*

Thouless, R.H. & Brier, R.M. (1970). The stacking effect and methods of correcting for it. *Journal of Parapsychology, 34, 124-128* [On-line] Available : *http://cabinet.auriol.free.fr/Documents/stacking-thouless.htm*

Warcollier, R. (1926). La télépathie expérimentale (travaux de groupe), *Revue Métapsychique, 4, Juillet-Aout, Alcan.*

Watt, C. (2003), unpublished discussion *at EuroPA meeting, 14th-17th November 2003, Port-Royal, France*

Softwares: Turbo Pascal v.7, Ms-Dos, Excel, S-Plus (MathSoft) and SAS.

Address for correspondence: Bernard M. Auriol, Institut Métapsychique International, 51, rue de l'Aqueduc, 75010, Paris, France. E-mail: auriol@free.fr

APPENDIX

Age	**Fischer's test statistics**	**P-value**
Date of birth	2.10	0.23

Lüscher's test	**Fischer's test statistics**	**P-value**
Blue	-0.69	0.49
Green	0.86	0.39
Red	-1.42	0.16
Yellow	-0.33	0.74
Black	1.16	0.25
Grey	-0.35	0.73
Brown	0.26	0.79
Violet	0.59	0.55

Sheep/goat questionnaire	**Fischer's test statistics**	**P-value**
telepathy	0.07	0.77
Clairvoyance	0.07	0.77
Precognition	1.02	0.37
Psychokinesis	0.08	0.75

Altered States of Consciousness	**Fischer's Test Statistics**	**P-value**
ASC practice	0.24	0.81

Table 2: Individual characteristics and hitting rate

Quartile of "vote strength"	**Number of trials**	**Hit Rate** (expected => 0.50)
1	323	0.50
2	590	0.49
3	489	0.49
4	481	0.51

Table 6: Collective hit rate for the two pictures protocol according to the quartile of "vote strength"

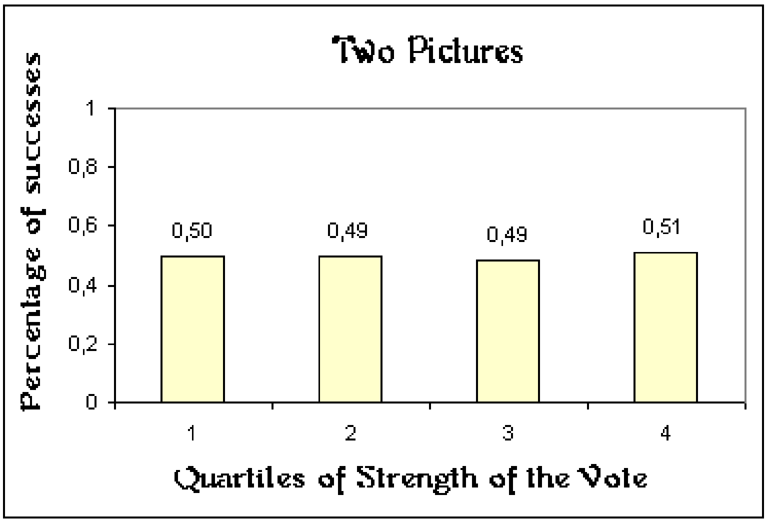


Fig. 8: Collective hit rate for the two pictures protocol according to the quartile of "vote strength"

Quartile of "Vote strength"	Number of trials	Hit Rate (expected => 0.33)
1	1871	0.32
2	1840	0.33
3	1833	0.33
4	1894	0.33

Table 7: Collective hit rate for the three words protocol according to the quartile of "vote strength"

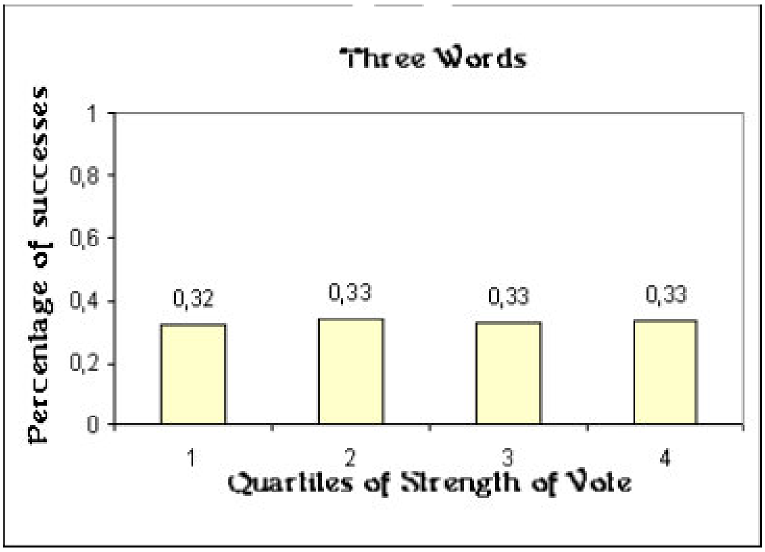


Fig. 9: Collective hit rate for the three words protocol according to the quartile of "vote strength"

Quartile of "Vote Strength"	Number of trials	Hit Rate (expected => 0.20)
1	2,148	0.19
2	1,308	0.21
3	1,632	0.21
4	1,857	0.21

Table 8: Collective hit rate for the five words protocol according to the quartile of "vote strength"

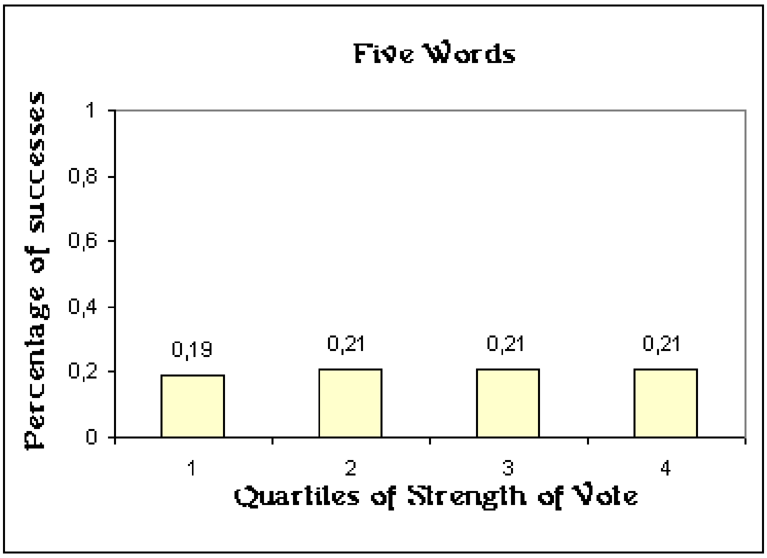


Fig. 10: Collective hit rate for the five words protocol according to the quartile of "vote strength"

	Names of variables	**Abbrev.**
Qual. Variables	The transmitting group has had an instruction	Instruction
	The participants (agents and percipients together) have chosen the list of potential targets	Choice
	There was a group feedback	G feedback
	There was an individual feedback	I feedback
	Maximum time for the vote = 20 seconds	20Max
	Maximum time for the vote = 60 seconds	60Max
	Maximum time for the vote = 120 seconds	120Max
	The time left to vote was shorted when the half of the percipients had voted	TimeLeft
Quant. Var.	Ratio : number of agents / number of participants	A/(A+P)
	Ratio : number of agents / number of percipients	A/P

Table 11: Names of variables and abbreviations

		Genuine data	First permutation	Second permutation	Third permutation	Fourth permutation	Fifth permutation
	Variable	*Estimate*	*Estimate*	*Estimate*	*Estimate*	*Estimate*	*Estimate*
	Intercept	+ 0.92	+ 0.91	+ 0.82	+ 0.78	+ 0.78	+ 0.93
Qual. Variables	Instruction	+ 0.10	+ 0.06			+ 0.10	
	Choice	- 0.29	- 0.32	- 0.30	- 0.32	- 0.30	- 0.32
	G feedback	- 0.11	- 0.08	- 0.09	- 0.11	- 0.10	- 0.06
	I feedback	+ 0.04		+ 0.06			
	20Max	+ 0.13	+ 0.20	+ 0.19	+ 0.21	+ 0.19	+ 0.15
	60Max	+ 0.19	+ 0.21	+ 0.24			+0.17
	120Max	+ 1.08	+ 1.04	+ 1.05			+ 0.98
	TimeLeft	+ 0.07	+ 0.13	+ 0.07	+ 0.12	+ 0.09	+ 0.13
Quant.	A/(A+P)	- 1.06	- 1.31	- 0.81	- 1.16	- 1.50	- 1.58
	A/P	+ 0.86	+ 1.02	+ 0.82	+ 0.93	+ 1.13	+ 1.13

Table 12: Analysis of the collective trials significantly higher than expected.

	Genuine data	**First permutation**	**Second permutation**	**Third permutation**	**Fourth permutation**	**Fifth permutation**
Variable	*Estimate*	*Estimate*	*Estimate*	*Estimate*	*Estimate*	*Estimate*
Resgr	- 1.48	- 1.49	- 1.32	- 1.86	- 1.34	- 1.39
Emesgr	+ 0.59	+ 0.38	+ 0.36	+ 0.35	+ 0.41	+ 0.40
R^2	0.11	0.11	0.07	0.13	0.09	0.09

Table 13: Sociometric bonds and stacking effect (trials significantly higher than expected p <.05)

CLASSIFICATION OF PHENOMENA OF PARAPSYCHOLOGY FROM THE POINT OF VIEW OF PHYSICS. SUPERFLUID VACUUM

Liudmila B. Boldyreva[1] & Nina B. Sotina[2]
[1]*State University of Management, Moscow*
[2]*Moscow State University*

ABSTRACT

From the point of view of physics all parapsychology-related phenomena can be divided into four categories: phenomena which contradict the modern physical knowledge obtained both theoretically and experimentally; phenomena which do not directly contradict the contemporary scientific knowledge but no explanation of which is seen at present; phenomena which have explanations on the basis of existing scientific paradigm; phenomena which allow for a physical interpretation on the basis of new concepts of physics that are consistent with experimental data.

Precognition can be placed in the first category as being inconsistent with the "arrow of time" (irreversibility), the notion introduced for fundamental description of nature. Note that precognition should not be confused with ordinary prediction, or a certain approximation into the future based both on the personal experience, knowledge of the expected course of events, and deeper sources of information feeding our intuition, such as telepathy. Telepathy refers to the second category (we look upon the telepathy from he point of view of physics and seek for the mechanism through which the sender selects the specific recipient from a number of recipients and the mechanism through which information propagates through the physical vacuum). Remote dowsing performed with the use of geographical maps, determination of the location of a person by photo or handwriting refer as well to this category. The third category includes some phenomena of psychokinesis where influence of an operator on a distant object can be explained by acoustic, diffusive or electromagnetic processes.

Among the phenomena of the fourth category are, first of all, the phenomena of psychokinesis which cannot be explained by thermal, acoustic or electromagnetic processes. These are the phenomena where the sensitive's influence is independent of the presence of electromagnetic screens, it features 'selectivity' and is also independent of the distance (up to 2000 km in the experiments). To explain such phenomena is to find a new physical mechanism underlying the effect with the above properties. We developed a model of physical vacuum as a superfluid consisting of pairs of unlikely charged microparticles with spins. (A superfluid consisting of pairs of fermions with zero total electric charge of pairs is known at present. It is ^3He). According to the model, a quantum object creates in vacuum certain structures which can interact with each other. This interaction may have the property of 'selectivity', it does not depend on the distance between the interacting structures and is performed 'without participation' of photons. In the framework of the model of the superfluid vacuum, many phenomena of psychokinesis can be interpreted as an interaction between the structures created in vacuum by the sensitive and the object being influenced. Ascribing the properties of superfluid to the physical vacuum allows one to work out recommendations on setting up experiments in the area of parapsychology (first of all, connected with shielding of the sensitive's influence). The processes in superfluids are theoretically dissipativeless. Thus, while describing a number of parapsychology-related phenomena, the vacuum with the superfluid properties can be treated as an information field.

INTRODUCTION

In this article a classification of phenomena of parapsychology from the point of view of physics is given. From this point of view all parapsychology-related phenomena can be divided into four categories: phenomena which contradict the modern physical knowledge obtained both theoretically and experimentally; phenomena which do not directly contradict the contemporary scientific knowledge but no explanation of which is seen at present; phenomena which have explanations on the basis of existing scientific paradigm; phenomena which allow for a physical interpretation on the basis of new concepts of physics that are consistent with experimental data.

Among the phenomena of the fourth category are, first of all, the phenomena of psychokinesis which cannot be explained by thermal, acoustic or electromagnetic processes. These are the phenomena where the where the operator's influence (the term *operator* will be used hereinafter for denoting the person exerting mental influence on an object: an instrument or organism) is independent of the presence of electromagnetic screens, features 'selectivity' and is also independent of the distance (the distance between the operator and the object influenced has been as great as 2000 km).

To explain such phenomena is to find a new physical mechanism underlying the effect with the above properties.

We developed a model of superfluid physical vacuum which is based largely on the properties of superfluid ^{3}He–B. According to the model, a quantum object creates in vacuum certain structures which can interact with each other. This interaction may have the property of 'selectivity', it does not depend on the distance between the interacting structures and is performed 'without participation' of photons.

In the framework of the model of the superfluid vacuum, many phenomena of psychokinesis can be interpreted as an interaction between the structures created in vacuum by the operator and the object being influenced. Ascribing the properties of superfluid to the vacuum allows one to work out recommendations on setting up experiments in the area of parapsychology, first of all, connected with shielding of the operator's influence. According to the model of superfluid vacuum, such shielding can be done by means of photons or rotating electric fields.

CLASSIFICATION

From the point of view of physics, all parapsychology-related phenomena can be divided into four categories.

The first category

Phenomena which contradict the modern physical knowledge obtained both theoretically and experimentally. Precognition (perception of some future situation in detail) is an example of such phenomena.

The rapid development of nonlinear physics and thermodynamics of nonequilibrium processes, which took place during the past decades, has revealed the necessity of introducing the "arrow of time" (irreversibility) for a fundamental description of nature (Prigogine & Stengers, 1997). It became clear that even in classical mechanics, while solving a specific task, one may face an unpredictable chaotic behavior of the coordinates and momenta which characterize the dynamic system state.

Various cases of successful fortune-telling are often cited to support the existence of precognition. Note that precognition should not be confused with ordinary prediction, or a certain approximation into the future based both on the personal experience, knowledge of the expected course of events, and deeper sources of information feeding our intuition, such as telepathy. Besides, the effect of self-suggestion based on one's belief in the truth of the prediction should not be ruled out.

The second category

Phenomena which do not directly contradict the contemporary scientific knowledge but no explanation of which is seen at present. An example of the phenomena of this kind is telepathy. Speaking of telepathy, we do not mean physiological, biophysical or psychological aspects of the problem, i.e. we are not interested in the processes taking place in the sender or receiver when he or she sends or receives targets/impressions in an ESP test. There is a lot of techniques used in here, such as "ganzfeld" described by many authors (e.g. Bem, 1996). Making our classification, we look upon the telepathy from the point of view of physics and seek answers to the following questions: What is the mechanism through which the sender selects the specific recipient from a number of recipients? What is the mechanism through which information propagates through the physical vacuum?

Remote dowsing performed with the use of geographical maps, determination of the location of a person by photo or handwriting refer as well to this category. Some examples are given below.

- Wolf Messing (Messing, 1990), widely known owing to his experiments on telepathy, described in his book *I am a Telepathist* how he got information about a young man whose photo had been given to him by the young man's sister.

> I am looking at the photo of the poor woman's brother... And all of a sudden I see him as if he got off the photograph. He looks younger; in a smart suit. I say, 'There is nothing to worry about, miss. Your brother is alive. He has had hard times, but now he is doing well. You will receive a letter from him in thirteen days, today inclusive.' It became the talk of the town. The information reached the journalists. When the thirteenth day forecasted by myself came, it appeared that the journalists of almost all Polish newspapers had gathered at that small town. The letter from far-off Philadelphia was delivered with the evening train. The fact received full coverage in the Polish newspapers.

- Wolf Messing could acquire information about a person by his or her handwriting. In the above-mentioned book he wrote (Messing, 1990), "I have taken a sheet of paper with a text... I see that the hand which wrote the text is dead." (It was established later that Messing's 'diagnosis' was correct).
- Certified biolocation engineer E.G. Bondarenko determined the location of oil-bearing layers using a map. In a control experiment, with the layers whose boundaries had been specified earlier, the objects were determined correctly in 10 out of 12 cases (Boldyreva & Sotina, 2002). (Note that people who knew the position of the layers were not present during the experiment.)

Note that a few researchers tend to explain some of these phenomena by quantum nonlocality. (Quantum nonlocality is a phenomenon where the results of quantum measurements conducted in two or greater number of points which are far-off from each other appear to be mutually dependent.) The idea of existence of quantum nonlocality emerged from the theory of composite states and the experiments with quantum-correlated elementary particles; the quantum-correlated particles have the following property: once having been described by the same wave function, their behavior in future (described by laws of the probability theory) will be mutually dependent even if they are spatially separated.

However, various explanations of quantum nonlocality have been generally considered arguable so far. There are several ways of "explanation" of quantum correlations: 1) the Copenhagen interpretation according to which nothing can be said about some properties of elementary particles before the measurement; 2) the existence of interactions which propagate at superluminal rates, 3) the introduction of negative probabilities (the rejection of "arrow of time".) Nevertheless, how can any of the above ideas serve as a basis for explanation of the following phenomenon: an operator telling by a person's photo whether the person is alive or dead?

The third category

Phenomena which have explanations on the basis of existing scientific paradigm. For example, some phenomena of psychokinesis[1] where influence of an operator on a distant object can be explained by acoustic, diffusive or electromagnetic processes. Below are given some examples of psychokinesis of that kind (based on the experiments conducted in Russia.)

- While investigating the acoustic field of an operator, Dr. G.N. Dulnev has detected acoustic pulses of 0.01 sec width and 70 dB amplitude in the frequency range of 25–40000 Hz (Dulnev, 2000). At certain instances of the operator's influence session, the pulse duration decreased down to 5–7 10^{-5} sec and the amplitude increased up to 90 dB. It is known that noise with such an amplitude can create pressure of 0.1 g/cm^2; the pressure is suffice to produce a displacement of a light object provided there is a considerable reduction of the friction force. To check the acoustic hypothesis of contactless displacement of objects as a

[1] Psychokinesis refers not only to the visible displacement of remote objects under influence of the operator but also to operator's influencing various organisms, such as bacteria, plants, or small animals.

result of the operator's (mental) efforts, the object was screened using a shielding castle. Under the normal pressure of 760 mmHg the object was displaced at the operator's influence. When forevacuum of 10^{-3} mmHg had been created under the castle, the operator was unable to move the object.

- A team of people who had been trained according to Dr Lee's technique (Lee & Ivanova, 1975) were able to rotate a non-metallic propeller placed under a glass castle from the distance of several centimeters. However, the effect became less as air was pumped off from the castle.

Note: If a human is able not only to radiate but also to detect ultrasound, then some cases of short-range psychokinesis could be explained by the effects of ultrasonography, e.g. recognition of Zener cards placed into light-tight paper bags.

- In 1985 N.D. Kolbun and V.E. Lobarev conducted experiments (a private communication, 1995) in which an operator acted positively on some biological objects at a distance of 10 cm. However, when an electromagnetic screen was used no positive effect was observed. The researchers concluded that the effect was of electromagnetic nature, which was confirmed by the fact that a similar effect took place when the same biological objects had been irradiated with electromagnetic fields (in the millimeter wavelength range).
- The research carried out by Dr. G.N. Dulnev (Dulnev, 2000) has shown that some people can generate intense thermal flows (radiating from the palm of a hand) of about 40 W/m^2. Such a radiation may underlie the positive effects produced by people influencing plants at distances of 4–7 cm.

Such phenomena are actively investigated now and, by tradition, are still believed to belong to parapsychology in spite of the fact that they are already tightly bound to classical scientific disciplines.

The fourth category

Phenomena which allow for a physical interpretation on the basis of new concepts of physics that are consistent with experimental data.

In our view the said phenomena are those which cannot be explained by thermal, acoustic or electromagnetic processes. As an illustration, we shall give below the results of experiments where operators influenced plants and microcalorimeters.

- At the beginning of the 1990's a team of Moscow researchers conducted experiments on the operator's influence on plant leaves (cucumber or corn were used) (Davydov et al., 1993). The electric response of a leaf to irradiation of the plant by light was measured, i.e. the voltage between the leaf and the root. It is known that the magnitude of the electric response depends upon the physiological condition of the plant and it changes under the influence of varying parameters of the environment. In the experiments, such a varying parameter was the influence of an operator, exercised during 3-4 min before the irradiation by light.

In the course of the experiments the measuring instruments and the plant were placed in a metal grounded chamber, and the operator exercised influence both in close vicinity of the plant and from an adjacent room (at a distance of 5 m).

122 experiments were conducted; among them 91 ones were control, in 15 experiments the operator aimed at reduction of the electric response, 16 at the increase of the response. It was established with 95 percent confidence that the operator could influence the physiological reaction rates of the plants, and the effect of the influence did not depend on the shielding of the experimental setup and the distance between the operator and the plant.

'Selectivity' of the operator's influence has been discovered. In 41 experiments there were 2 closely placed plants (within a distance of 40 cm) and the operator attempted to influence only one of them at a distance. It was shown with 95 percent confidence that the operator was able to exert selective influence. The effect of such selective influence did not depend on whether the operator was in the same room as the plants were, or in a different room.

- At the end of the 1980's, two Moscow scientists, Dr. G. K. Gurtovoy and Dr. A. G. Parkhomov, carried out experiments on the operator's influence on microcalorimeters (Gurtovoy & Parkhomov, 1992). The microcalorimeter is an instrument intended to measure weak thermal effects. It can measure the temperature difference (through measuring the thermistor's electrical resistance) of the order of 10^{-5} °C.

An analysis of the results of a great number of trials revealed the following characteristics of the operator's action at a distance.

There was 'selectivity' of the operator's influence: the operator was able to influence the microcalorimeter placed at a distance of some kilometers from the operator, and, at the same time, did not influence the instrument placed in an adjacent room.

Thus in the above experiments the operator's influence cannot be explained by thermal or acoustic factors due to the great distance between the operator and object being acted upon. Besides, the independence of the distance and presence of electromagnetic screens, as well as the 'selectivity' of the operator's influence rule out the electromagnetic hypothesis too. Thus, to explain the above effects, it is necessary to search for a new type of physical mechanism accounting for conveying the operator's influence. To this end we address the properties of physical vacuum.

PHYSICAL VACUUM

Before the special theory of relativity was introduced, space was assumed to be filled with a medium (ether) with the properties similar to a common "mechanical" medium. Since such a model of ether could not explain a number of experimental facts, the theory of special relativity discarded the ether as a physical medium.

According to a modern cosmological theory, the universe is filled with the physical vacuum which created an antigravitational field giving rise to the cosmological expansion acceleration (Chernin, 2001). Based on the general theory of relativity the cosmological theory presumes that the space vacuum, irrespective of the frame of reference, would universally and invariably have the constant density of energy (which should exceed the density of any other kind of space energy) and also the constant negative pressure. From the point of view of modern physics, the space vacuum is the same physical vacuum where interactions between elementary particles take place and which is present in the atomic physics as the state of field with minimum energy. However, the zero oscillations of vacuum energy, the polarization of vacuum in electric fields, the Lamb shift of spectral lines of atoms, the electron-positron pair production suggest that the microstructure of physical vacuum is complex and only under sufficiently great averaging scales it can be assumed that the pressure is universally and invariably constant and the energy density is constant.

In our works (Boldyreva & Sotina, 1992, 1999, 2002) we presented a model of vacuum as a superfluid consisting of pairs of unlikely charged microparticles with spins. The total spin of the pair equals zero. (A superfluid consisting of pairs of fermions with zero total electric charge of pairs is known at present. It is ^{3}He.) The model is substantiated well enough. The superfluid properties of vacuum (the zero viscosity while in motion) explain the observed dissipativeless motion of celestial bodies in space. The presence of electrically unlike microparticles allows one to describe the dielectric properties of the vacuum and the production of electrically charged elementary particles out of the vacuum.

Some researchers (Sinha & Sudurshan, 1978) have advanced earlier a model of vacuum as a superfluid consisting of pairs of unlikely charged particles with spins. In some papers, e.g. (Bauerle et al., 1996), an analogy is discussed between the phase transitions in the vacuum and those in condensed media, specifically, in superfluid ^{3}He-*B*. However, the four-dimensional relativistic formalism is preserved in all of the proposed models; in some of them the photon is considered a collective excitation of fermionic field, but the nature of the excitation has not been specified in them.

We develop the model of superfluid vacuum in the framework of the three-dimensional Euclidean space and time independent of the spatial coordinates. We extended the analogy between the properties of superfluid ^{3}He-*B* and those of the physical vacuum. Such properties of the superfluid ^{3}He-*B* vortices as quantization of the angular momentum, production of a core with altered inertial properties, electric polarization (Salomaa & Volovik, 1987) allowed us to suggest that the photon was a vortex structure in the superfluid physical vacuum.

Essentially, our model treats the superfluid vacuum as a luminiferous medium which, historically, might be best called the superfluid ether. No doubt that the introduction of such a medium arouses the well-known problems of the preferable frame of reference, which have been resolved by special relativity at the cost of discarding the ether. In our works, we represented the photon (in the pure state) as a complex entity with intrinsic degrees of freedom, which allowed us, with due account for the fundamental role of measurement in quantum physics, to advance new promising approaches to solving those problems.

All superfluids are able to sustain for a long time the structures (e.g. vortices) created in it. But owing to the fact that a pair of unlikely charged particles in superfluid ^{3}He has nonzero summary spin this superfluid has unique properties in this respect: stable spin structures may exist in it. The topological structures created in the fluid were studied in a number of fundamental works. In our model, any quantum object described by the Schrödinger equation (e.g. an electron in an atom) creates in the vacuum a spin structure of the uniformly precessing domain (UPD) type.

There are processes in superfluids that lead to the equalization of the order parameter. The order parameter gradient can be caused, for example, by a difference in orientation of spins of the fluid particles or a difference in the spin precession phases. The order parameter gradient may give rise to a superfluid spin current. At a certain magnitude of the spin current, a precession phase slip of the value of $2\pi n$ can occur. The following features of the process are worthy of notice: quantization of the phase slip value at which the phase slip takes place; 'selectivity' of the effect, namely, the phase slip takes place only at specific magnitudes of spin current (or phase gradient, which is the same); independence of the distance between the interacting structures, i.e. the UPD's.

Thus, according to the model of superfluid physical vacuum, which is based largely on the properties of superfluid ^{3}He-*B,* the quantum objects create in the vacuum certain structures which can interact between each other (for example, by means of spin current). This interaction has the properties:

- 'selectivity';
- does not depend on the distance between the interacting structures;
- is performed 'without participation' of photons.

DISCUSSION

In the framework of the model of the superfluid vacuum, many phenomena of psychokinesis which cannot be explained by thermal, acoustic or electromagnetic processes (in the classification given in this paper some such phenomena of psychokinesis were related to *the fourth category*) can be interpreted as an interaction between the structures created in the vacuum by the operator and the object being influenced (Boldyreva & Sotina, 2002).

Ascribing the superfluid properties to the vacuum allows one to work out recommendations on setting up experiments in the area of parapsychology.

-Shielding of the operator's influence with rotating electric fields. In our model of superfluid vacuum, rotating electric fields give rise to a precession of the particles constituting the vacuum. This should affect the interaction between the structures created in the vacuum by the quantum objects.

-Shielding of the operator's influence with a flow of photons. As was mentioned above, in our model of superfluid vacuum, photons are vortical structures: the precession of the particles constituting the vacuum takes place in those structures. This should, as in the previous case, affect the interaction between the structures created in the vacuum by the quantum objects.

-Affecting the 'selectivity' of the operator's influence (that is affecting the 'connection' between the operator and the object being influenced upon.) Our model of the physical vacuum presumes that the 'selectivity' is of frequency-related nature.

The processes in superfluids are theoretically dissipativeless. Thus, while describing a number of parapsychology-related phenomena, the vacuum with the superfluid properties can be treated as an information field.

References

Bauerle, C., Bunkov, Yu. M., Fisher, S. N., Godfrin, H. & Pickett, G. R. et al. (1996). Laboratory simulations of cosmic string formation in the early universe using superfluid *He. Nature, 382*, 332.-334.

Bem, D. J. (1996). Ganzfeld phenomena*. In G. Stein (Ed.), Encyclopedia of the paranormal*, 291-296. Buffalo, NY: Prometheus Books.

Boldyreva, L. B. & Sotina, N. B. (1992). *Superfluid Vacuum with Intrinsic Degrees of Freedom*. Physics Essays, **5,** 510.

Boldyreva, L. B. & Sotina, N. B. (1999*). A Theory of Light Without Special Relativity?* Moscow: Logos.

Boldyreva, L. B. & Sotina, N. B. (2002). *Physicists in parapsychology*. Moscow: Hatrol.

Boldyreva, L. B. & Sotina, N. B. (2002). Superfluid Vacuum as a Basis for Explanation of Some Phenomena of Parapsychology. *ISLIS, 20* (2), 543–548, Japan.

Boldyreva, L. B. & Sotina, N. B. (2002). The Possibility of Developing a Theory of Light Without Special Relativity. *Galilean Electrodynamics, 13* (6), 103-107.

Chernin, A. D. (2001). The Cosmic Vacuum. *Physics Uspekhi, 171*, 1153-1175 [in Russian].

Davydov, V. A., Dolin, Yu. A., Morozova, E. V., Shumov, D. E. (1993*). Studies of remote mental effect on plants with electric physiological records. The Protocol of Parapsychological Association, 36 Annual Convention,* Toronto, Canada.

Dulnev, G. N. (2000). *Energo-information exchange in nature*. St. Petersburg: IVA [in Russian].

Gurtovoy, G. K. & Parkhomov, A. G. Experimental study of operator's remote influence on physical and biological systems. *Parapsychology and Psychophysics, Vasiliev Fund of Parapsychology. 4,* 31-51, Moscow [in Russian].

Lee, A. G. & Ivanova, T. K. (1975). The development of techniques of effort synchronization in a team of operators in telekinesis experiments. *Parapsychology and psychophysics, 1*, 22-46 [in Russian].

Messing, W. (1990). *I am a Telepathist.* Moscow: South-West [in Russian].

Prigogine, I. & Stengers, I. (1997). *The End of Certainty; Time, Chaos and the New Laws of Nature.* New York: the Free Press.

Salomaa, M. M. & Volovik, G. E. (1987). Quantized vortices in superfluid ^{3}He. *Reviews of Modern Physics. 59*, 533.

Sinha, K .P. & Sudurshan, E. C. G. (1978). The Superfluid as a Source of All Interactions. *Foundations of Physics. 8***,** 823.

Address for correspondence: Liudmila B. Boldyreva,
State University of Management, Academica Anokhina
Ulitsa 5-1-61, 119571 Moscow, Russia.
E-mail: m_boldyrev@mtu-net.ru.

PSI TEST FEATS ACHIEVED ALONE AT HOME: DO THEY DISAPPEAR UNDER LAB CONTROL?

Suitbert Ertel

Georg-Elias-Müller-Institut für Psychologie, Göttingen

ABSTRACT

Extraordinary hit rates from multiple choice tests, obtained by participants alone in their homes, are ambiguous. On the one hand, their feats might reflect psi power manifesting itself better under informal home than under lab conditions. Yet hit excesses obtained without control might also be due to negligent or fraudulent conduct. One way out of this dilemma is to let participants complete psi tests at home and to invite high scorers thereafter to do additional runs under lab control. This strategy has been endorsed using N = 238 (sample I) and N = 47 (sample II) of student participants. Sample I (female 84%) completed the ball drawing test, version I. Table tennis balls are drawn from an opaque bag on which numbers 1 to 5 are written, each number on ten balls. Participants guess and draw numbers blind and record their guessed and drawn numbers (hit expectancy 20%). One test unit consists of six or eight runs comprising 60 trials each (total 360 or 480 trials). Participants shake the bag prior to each trial and put drawn balls back into the bag. Sixteen high scoring participants of sample I were also tested, using the same test, under lab control. Sample II (female 73%) completed the ball drawing test, version II. This test resembles test I except that green or red dots are sprinkled over the balls, participants guess numbers (five targets) and colours (two targets), the combined expectancy being 10%. Thirteen high scorers of sample II were also tested under lab control using a pearls drawing test where they draw one of five colours (no numbers, expectancy 20%). Hypotheses: 1. Hit rates of high scorers in home tests decline (due to less psi-conducive conditions under control and regression towards the mean). 2. Hit rates of high scorers under control score still significantly above chance (due to genuine psi which was also effective at home). Both hypotheses were confirmed with sample I and replicated with sample II. *Sample I* obtained an average ES = 0.369 (sd=.126) under home condition and an average ES = .122 (sd = .207) under control ($ES = Z/\sqrt{(N_{trials})}$), the difference is significant ($t_{(corr.smpls.)} = -3.02$, df = 15, p = .004, one-tailed). The significance of hit rates under control is $p = <10^{-15}$ by $Chi^2 = 721.7$, df = 16 ($Chi^2 = \Sigma(Z)(Z)$). *Sample II* obtained an average ES = .275 (sd= .162) under home condition and an average ES = .098 (sd = .153) under control, the difference is significant (t = -3.00, df = 12, p = .006, one-tailed). The significance of hit rates under control is $p = <10^{-15}$ by $Chi^2 = 223.9$, df = 13. Surprisingly, three participants obtained significantly *higher* hit rates under control compared with their home performance. The issue of fraud and bias loses relevance in view of such finding. It is recommended to introduce the "*first-home-then-lab-test*" strategy in parapsychological research on a broader scale. Once this strategy were generally applied, the widespread lamenting of psi researchers and their critics about tiny and elusive experimental psi effects might come to an end.

INTRODUCTION

Out of necessity, not out of choice (funds and assistants were lacking) I have been conducting experimental research into the paranormal ignoring, with qualms, the community's methodological demand #1: Fraud must be ruled out in the first place (Milton, 1996). I invited students to complete psi tests at home. The students, participating at this project out of curiosity and/or with intent to obtain obligatory credit points, were instructed in class, took the material home, ran the required 360 – 480 trials within 8 days (total testing time 1 ½ hours), returned their record sheets and waited for feedback from the experimenter who analyzed the data individually and who informed each participant in detail, by phone or email, about his or her results. Five annual cohorts of student beginners (total N = 238, female 84%) were tested using the ball drawing test (version I): From an opaque bag, participants drew numbers 1 to 5 written on table tennis balls (each ball one number) after having guessed at each trial which number they would draw next. Another ball test, somewhat more complex (version II), was used for cohort #6 (N = 47, female 73%): Each target ball carried numbers 1 to 5 plus green or red dots, the participants guessed numbers and colors. The last student cohort # 7 (N = 48, female 86%) had pearls to draw from a box, five different colors (no numbers) served here as targets. The median age of the total is 23.2 years, age differences between cohorts are negligible.

The ball drawing test yielded considerable hit surpluses from chance (see Table 1 and Figure 1, open circles). Hit rates declined with cohorts 2000–2001, but cohort 2002 regained the level of cohorts 1998/1999. Each cohort's hit surplus was highly significant. The results of ball test II exceeded those of ball test I. An analysis of the pearls drawing data is more complex[1], the result may be condensed by saying that deviations from chance were as large as those of the best sample of ball test I participants (i.e., of cohort 2002).

Table 1: Results of home tests by student beginners across seven cohorts (columns). Balls tests I and II and pearls test (see row #02). Rows ## 08 - 11 provide effect size and inferential results.

01	Cohort no.	1	2	3	4	5	6	7	
02	Cohort year	1998	1999	2000	2001	2002	2003	2004	
03	Test	Balls I	BallsI	Balls I	Balls I	Balls I	Balls II	Pearls	
04	Targets	Numbers	Numbers	Numbers	Numbers	Numbers	Ns. & colours	Colours	
05	N Ss	57	38	56	36	44	47	48	
06	Trials	13,680	9,120	20,160	12,960	15,840	22,560	70,121	
07	Hits	3,021	2,011	4,316	2,759	3,539	2,620	14,400	
08	% exp	20.00	20.00	20.00	20.00	20.00	10.00	20.00	
09	% obs	22.08	22.05	21.41	21.29	22.34	11.61	20.54	
10	% surplus	10.40	10.25	7.05	6.45	11.70	16.10	2.70	
11	ES=Z/√Trials	0.052	0.051	0.035	0.033	0.059	0.054	0.013	0.029
12	$Z_{binomial}$	6.1	4.9	5.0	3.7	7.4	8.1	3.54*	7.6
11	p	10^{-8}	10^{-6}	10^{-6}	10^{-4}	10^{-13}	$<10^{-15}$	0.0002	10^{-13}

Note: % exp = expected hit percentage. % obs = observed hit percentage. % surplus = percent of hits above expectancy. Example: 22 hits, 20 expected = 10% surplus. Expectancy is different in column, 6, see row 06. Col 7 left = Hit average. Col. 7 right = Combined hit avg. and differential color effect. * $Z_{Poisson}$ = 3.50.

Yet all such results might be doubted because the observed deviations from chance might be brought about by negligent participants who might gain leaking sensory information or who might deceive themselves by inadvertently recording wishful hits while actually having misses. Without control, even intentional deception is feasible. Unlike ordinary psi test data where bias and fraud are meticulously ruled out, the present data base is not safe.

[1] The significance of hit rate deviations of the pearls test (hit counts fixed, trial counts variable) should generally be obtained by applying the Poisson test (Timm, 1968, 1994). For large trial counts, however, as for the present total of 70,121 trials, Binomial test results are numerically almost equal to Poisson test results.

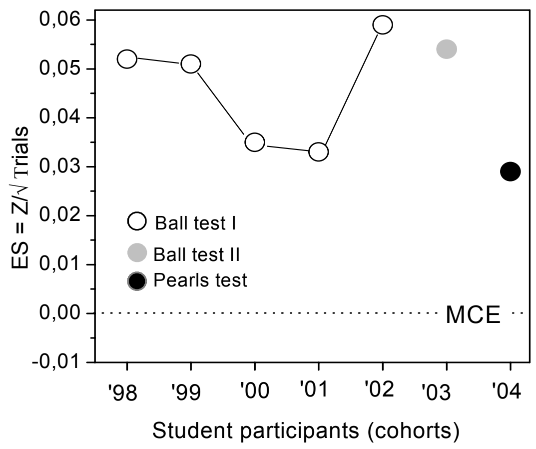


Figure 1: Effect size of hits for seven cohorts of student beginners, cohorts 1998-2002 took the ball drawing test I, cohort 2003 the ball drawing test II, and 2004 took the pearls drawing test.

However, unlike aseptically clean data where bias and fraud *cannot* have occurred, the present data are suitable for investigating and possibly finding out, post hoc, whether or not bias, e.g., by sensory leakage or memory help or fraud *have actually* occurred.

The issue of sensory leakage may be solved, above all, by checking for signs of ordinary learning. Sensory leakage cannot take effect on first trials, tactile or temperature information from balls and pearls must be perceived, stored and applied so as to become helpful subsequently for better hitting. The same applies to wishful errors at record taking that might eventually increase due to gradually fading attention. But the students' hit rates, those of high and low scorers alike, exceed chance expectancy already with first trials, there is no indication of increasing performance over time (Ertel, in print). [2]

The issue of fraudulent misconduct may be addressed with blunter or sharper means. A blunter test, nevertheless indispensable in professional criminal practice, as is well known, is to unravel motives.[3] Student participants do *not have to* complete these tests, most participants take the opportunity to find out whether they score well. They cannot satisfy such curiosity by manipulating their records.[4] Self report information by students (14-15 ys) about cheating in class (Anderman et al., 1998) shows that 39% admitted having cheated occasionally when taking exams or completing homework. Extrinsic motivation (getting grades etc.) and expected reward (escaping effort) turned out to be the prime motivation. The sample of students used in the present ESP studies are adults, no grades or other extrinsic reward could be expected for good hit rates, rather intrinsic motivation was manifested in responses to a post-experimental questionnaire of the pearls test sample (N = 48): Among 28 items addressing attitudinal and affective reactions to this test the first eight of an average rank order of affirmation are (1) accuracy, (2) interest, (3) curiosity, (4) joy when getting hits,

[2] An earlier general objection to ESP methods raised by Goodfellow (1938), Feller (1940), Brown (1953), Pöppel (1967), and Gatlin (1979) and revived by Brugger&Taylor (2003) has been scrutinized elsewhere (Ertel, submitted). This objection drawing upon patterns of calls and hypothetical patterns of targets matching with call patterns did not find support by various empirical tests using the author's ball test data.

[3] Hansen (1990) advocates considering not only the possibility of cheating, but the probability of its occurrence in the first place (*"…One must first assess the likelihood of fraud being attempted… [Considering probability,] even critic C. E. M. Hansel (1966) has stated: 'It is unlikely that more than a small number of experiments on ESP are affected by cheating'* " (Hansel, 1990, p. 32).

[4] Self report information by students (14-15 ys) about cheating in class (Anderman et al., 1998) shows that 39% admitted having cheated occasionally when taking exams or completing homework. Extrinsic motivation (getting grades etc.) and expected reward (less effort) turned out to be the prime motivation. The sample of students used in the author's ESP studies are adults, no grades or other extrinsic reward could be expected for good hit rates, rather intrinsic motivation was manifested. Moreover, regarding the proportion of cheaters at exams it should be considered that occasional cheaters do not cheat at every exam, the average proportion of cheaters at individual exams is much less than 39%.

(5) conscientiousness, (6) eagerness, (7) doubt of extrasensory perception, (8) fun. The eight most frequently negated items are (28) concern, (27) competition with fellow students, (26) lack of concentration, (25) fears, (24) strange self perceptions, (23) confusion, (22) thoughts about possible cheating, (21) statistical calculations.

Moreover, the participants had been informed by instruction that high scorers at home would be invited to subsequent lab tests under control. If a student under home condition were tempted, say, by deviant motives, to deceive the experimenter, he or she would have to consider that deception might eventually be discovered, a risky undertaking for fresh participants of academic courses. [5]

A sharper way of addressing the fraud issue is to actually invite good home scorers to complete additional runs in the lab. The present paper gives an account of this endeavor. Two questions need to be answered: 1. Do high scorers under home test conditions produce random hit rates under control? If they do, hit surpluses obtained at home may be explained as due to bias, fraud or other such "normal" factors, psi hypotheses would be dispensable. 2. If hit rates under control are lower than without control, but *not* random, two additional questions arise. 2.1 Are the remaining hit surpluses still strong enough to allow for assuming paranormal factors? 2.2 If they are strong enough, why do hit rates under control decline?

METHODS

Materials

Ball drawing test I (numbers only): An opaque bag containing 50 table tennis balls. The numbers 1 to 5 are written on them, each ball carries one number. Ten balls are used for each number. The face used is inaccessible to kinesthetic perception. Record forms for the home tests are provided where participants have to fill in their guessed and drawn numbers.

Ball drawing test II (numbers and colors): This test differs from the foregoing test only in that each ball is sprinkled, in addition, with either red or green dots. Red and green colors are used equally often. The record forms provide blank space to be filled in with guessed and drawn numbers and colors.

Pearls drawing test: A box containing 1,500 little handicraft pearls (diameter: 4 mm) whose colors are red, green, blue, yellow or white, 300 pearls for each color. Record forms provide blanks to be filled in with frequencies of blind draws which are made to pick a desired color (see procedure).

Procedure

Ball drawing tests: Each trial starts off with first putting down, on the record sheet, a guessed number (or number plus color, respectively) that the participant supposes to draw on the next trial. He then shakes the bag, closes his eyes, reaches into the bag, picks and draws a ball, checks the actually drawn number (and color) and writes it (them) down. He puts the ball back into the bag. Each run requires 60 trials, one ordinary test series requires six runs.

For the ball drawing tests, the number of trials is fixed while the number of hits is open to variation as is common with multiple choice procedures. For the pearls drawing test, the number of *hits* is fixed while the number of *trials* is open to variation, as follows:

The pearls drawing test: The record form is made up of rows with 10 blanks each. The participant fills them in with *frequencies of blind draws which were needed until a pearl with the desired target color was picked.* Prior to starting off with one row of blanks, the participant decides the target color for that row. Supposing he chooses "red" and picks a red pearl at his first trial, he fills in "1", if another pick of a red pearl requires, say,

[5] The idea of "reducing experimenter control in a study of special subjects" is not new. Bierman & Gerding (1992) advocate more relaxed ("sloppy") testing conditions in case of "special" subjects, control might be "reduced to theoretically relevant" factors. "It is argued that strong emphasis on this issue [control over all factors] has hindered progress in parapsychological research." "...One sometimes gets the impression that the demand for control is not determined by legitimate concerns about methodological rigor, but instead by the conservative attitude of so-called skeptics" (Gerding & Bierman, 1997, p. 2).

10 draws, he fills in "10" in the second blank etc. Each draw of a pearl is preceded by stirring the pearls in the box with stretched five fingers. The pearls are drawn with eyes closed. Drawn pearls are put back into the box. One sub-series consists of 10 rows, thus 100 hits are required for one sub-series, MCE being 500 draws. Three sub-series, i.e. 300 hits, MCE 1,500 draws, are needed for one complete unit. The participants are told to use the five colors row-wise as targets within one sub-series equally often.

Testing under lab control

The participant completes the test sitting opposite the experimenter. The participant's guesses and draws are filled in the blanks by the experimenter who also observes the participant's behavior. Attempts to peep into the box or other such fraudulent actions would hardly go unnoticed, apparently they have not occurred.

Participants and experimenters

The ball drawing test I (numbers) was completed by 238 students, the ball drawing test II (numbers and colors) by 47 students. From ball test I participants 16 high scorers completed the *same test* under control. Among them were 5 non-students, two boys (8 and 12 years) and three adults who had shown high hit rates at home. Of the balls test II sample, students only, 13 high scorers completed the *pearls test* under control (second control sample). High scorers among students who completed the pearls test at home (N = 49, the most recent cohort), have not (yet) been tested under control. The author acted as experimenter for 15 of 16 participants of the first control sample and for 8 of 13 participants of the second control sample. Six control experiments were conducted with student assistants serving as experimenters.[6]

Data analysis

Mean chance expectancies (MCE) are as follows: 0.2 for the balls test I, 0.1 for the balls test II, and 0.2 for the pearls test (see Table 1, row #06). The ball test data are subjected to the Binomial test. The pearls test data (hit count fixed, trial count variable) might be subjected to the Poisson test (as advocated by Timm 1968 and 1994). But there are grounds to apply the Binomial test consistently.[7] Each individual Binomial Z value is transformed into effect size $ES = Z/\sqrt{T}$ [T = Number of trials]. The significance for samples of individuals is obtained by $Chi^2 = \Sigma Z^2$, df = N [N = number of individuals].

RESULTS

Figure 2 shows results of sample #1 participants, i.e. of those who completed ball test I under home and lab conditions. Effect size measures of hit rates obtained under home (left) and control condition (right) are plotted. Observed effect size of the total (lower dot, left) is plotted for comparison. Necessarily, the difference between the effect size of selected high scoring participants (N = 16, full circle upper left) and that of the total (N = 238, full circle lower left) is large. More important, the effect size under control (right) is remarkably lower than the effect size obtained under home condition (left). However, even though some participants obtained hit rates under control at chance level only (N = 6, = 38% of 16), hit rates of the high scorers' total remained highly significant under that condition ($Chi^2 = 721.7$, df = 16, $p < 10^{-15}$).

---

[6] Three student experimenters had completed the test under home and lab conditions earlier (control by SE). They tested two students each.

[7] The pearls' test condition is atypical for applying the Poisson procedure which is applicable in cases of very rare events (p<.20) and when the sample size is unknown. The pearl's test probability is .20 and the number of hits is fixed (N = 300), hence the range of necessary trials is also *roughly* known, empirically (1100 – 1600) (see also Footnote 3).

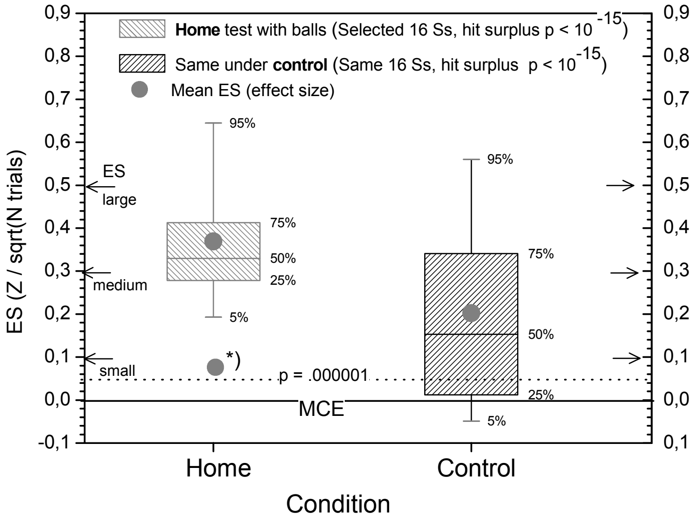


Figure 2: Sample #1 results (N = 16). Effect size from ball test I obtained at home (left) and under control in the lab (right). The lower left full circle, marked *), represents the observed effect size of the total from which sample #1 had been selected (19,442 hits, exp. 0.2, by 85,200 trials of 238 Ss). The dotted p-line is based on the total.[8]

Figure 3 shows results of sample #2 participants (N = 13), i.e., of those who completed ball test II at home (left) and the pearls test under lab conditions (right). The mean hit rate of high scorers of sample #2 is less pronounced than that of sample #1 (Figure 2) because the selection criteria for sample #2 were less strict. More important, the average effect size for hit rates of sample #2 participants decreased under lab control as was the case with sample #1 participants. The home-lab difference between hit proportions is highly significant ($Z = 6.36$, $p = 10^{-10}$). Under control, four of 13 participants obtained hit rates at chance level (= 31% of 13), nevertheless, the hits total of high scorers remained highly significant under that condition ($Chi^2 = 223.9$, $df = 13$, $p < 10^{-15}$).

[8] Box plots show means (here circles) medians (here solid lines within boxes), 25% and 75% centile values (lower and upper boundaries of the boxes) and 5% and 95% centile values (lower and upper ends of vertical lines). The end of the lower vertical (5%) of controls (right) extends below ES = 0 because a small number of participants with significant hit surpluses at home obtained hit scores below expectancy under control.

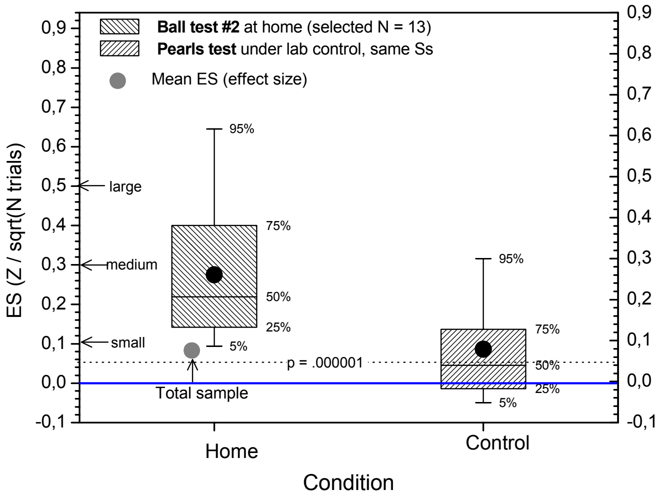


Figure 3: Sample #2 results (N = 13). Effect size from ball test II obtained at home (left) and from the pearls test under control (right). The lower left full circle, marked *), represents the observed effect size of the total from which sample #2 had been selected (2,620 hits, exp. 0.1, by 22,560 trials of 47 Ss). The dotted p-line is based on the total.

For the above data analysis individual scores were taken as units, the box plots of Figures 1 and 2 are based on them. Alternatively, hits may be summed across participants ignoring individual hit rate variation. This procedure has been applied with Table 1 and is taken up again with Table 2 which summarizes results for the two control samples (see Table 2).

Table 2: Results of samples #1 and #2 obtained under home condition (ball-drawing I for sample I, ball drawing II for sample II) and under control condition (ball drawing I for sample I) and pearls drawing test (for sample II).

	Sample	I		II	
	N participants	16		13	
01	Condition	*Home*	Control	*Home*	Control
02	Test	Balls I	Balls I	Balls II	Pearls
04	Trials	11,040	9,360	7,800	15,244
05	Hits	3,996	2,628	2,033	3,400
06	% exp	20	20	20	20
07	% obs	36.20	28.08	26.06	22.30
08	% surplus	81.00	40.40	30.03	11.50
09	ES=Z/√Trials	0.405	0.201	0.152	0.053
10	$Z_{Binomial}$	42.53	19.5	13.38	7.1 *
11	p	$<10^{-15}$	$<10^{-15}$	$<10^{-15}$	10^{-12}

Note: * $Z_{Poisson}$ = 6.73
For other clarifications see Notes of Table 1

This mode of analysis confirms that hit rates of the two samples tested under control are extraordinary.

DISCUSSION

We set out by asking: "(1). Do high scorers under home test conditions produce random hit rates under control? If they do, hit surpluses obtained at home may be explained as due to bias, fraud or other such 'normal' means, psi hypotheses would be dispensable." The results show that high scorers under home conditions, when tested under control, did not produce hit rates at random. Paranormal abilities must have been effective when they completed the tests under control, hence it would be absurd to surmise that such ability was not effective when control was absent.

We also asked: "2. If hit rates under control are lower than without control, but not random, two additional questions arise. 2.1 Are the remaining hit surpluses still strong enough to allow for assuming paranormal factors? 2.2 If they are strong enough, why do hit rates under control decline?" The results presented in Figures 2 and 3 showed that hit rates of high home scorers, even though lower on average under control, were in fact strong, hence no other than paranormal factors (such as statistical flukes etc.) need consideration. The only issue of uncertainty is why hit rates declined in the lab. Two main factors must be considered.

One straightforward factor is regression towards the mean (RTM). For simple stochastic reasons, exceptional hit rates which may be due to or boosted by chance fluctuation tend to be followed by less outstanding performance, hence the observed change in hit rates from *much* to *less* might at least partially be explained by RTM.

More important, a decline by control might be due to the lessening of psi-conducive conditions. One of the participants of sample #1 expressed concern:" *I am afraid that if my score comes out random under control, you might think that my home records were fudged.*" Other lab-tested participants might have felt similarly. Researchers of the paranormal regard stress, tension, and anxiety as psi-inhibiting factors. An undisputed requirement for psi research is to provide "social ambience" in order to make fearful participants feel more comfortable. "*Interaction with the subject should be calm, friendly, positive, and unhurried*" (Reinsel, 1999). Yet stress-free conditions as are generally experienced at home can hardly be brought about in the lab by kind words with offering cakes and drinks.

The issue of whether situational conditions affect psi performance, however, is thornier. We noticed that conducting psi tests in the lab did not have detrimental effects for all participants alike. Curiously, for three high scorers at home, for Ahmed, Amelie, and Silke, hit rates under control even increased immensely as is shown in Figure 4 and Table 3. This observation can hardly be explained in terms of diminishing psi-conducive conditions since for them effects by control cannot have surpassed those of a relaxed home atmosphere. The conclusion remains ambiguous: The presence of experimenters exerts either hit diminishing (more often) or hit enhancing effects (less often) on psi-gifted participants. In any event, the occurrence in three cases of psi-enhancing effects by control shows that fraud and bias and the like are largely irrelevant when it comes to explain hit score differences between controlled and uncontrolled conditions.

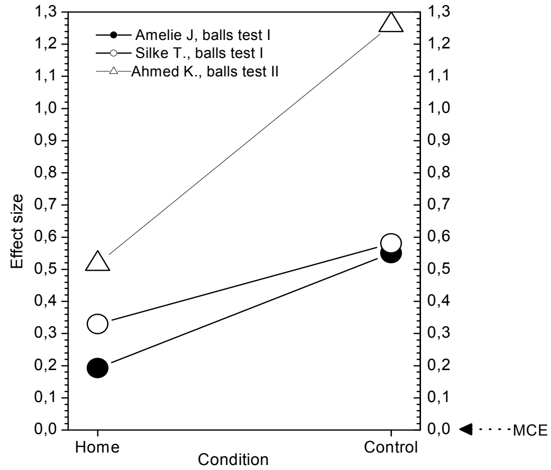


Figure 4: Effect size of hit rates obtained by three participants under home (left) and lab control (right) conditions. Curiously, their hit rates increase under lab control.

Table 3: Original and derived data of three participants whose hit rates increased significantly and extraordinarily under control. Significance tests for Binomial Z's and fir differences of proportions.

	Participant	Amelie J.		Silke T.		Ahmed K.	
01		Home	Control	Home	Control	Home	Control
02	Test	Balls I	Balls I	Balls I	Balls I	Balls II	Balls II
03	Trials	240	960	360	480	480	480
04	Hits	67	436	120	204	123	230
05	% exp	20.0	20.0	20.0	20.0	10.0	10.0
06	% obs	27.9	45.4	33.3	42.5	25.6	47.9
07	% surplus	39.5	127.0	66.5	112.5	156.0	279.0
08	ES=Z/√Trials	0,193	0,550	0,330	0,580	0.517	1.260
09	$Z_{binomial}$	7.83	29.5	6,26	12.27	11.33	27.6
10	P	10^{-14}	$<10^{-40}$	10^{-9}	10^{-34}	10^{-29}	$<10^{-40}$
11	Z (of difference)	4.91		2.70		7.16	
12	p	10^{-6}		.0035		10^{-12}	

Note: If clarifications are needed, see Notes of Table 1

It is not claimed here that negligent or fraudulent actions cannot have occurred under home test conditions in any single case. The claim merely is that if misconduct occurred at all, its contribution to the overall deviation of hit rates from chance is negligible.

One final comment on the difference between samples #1 and #2: Hit rates of sample #2 were less pronounced than those of sample #1, under both, home and lab conditions (see Figures 2 and 3). The difference in home test performance is simply due to applying different selection criteria, they were less strict for sample #2 than for sample #1. Part of the difference under the lab condition may therefore be explained by less pronounced psi abilities of sample #2 participants. In addition, the test used in the lab for sample #2 high scorers (the pearls drawing test) differed from the test that they had used at home (the ball drawing test II). Procedural changes effectuating some loss of familiarity with conditions might have an unfavorable effect on performance. In short, comparing the two sample's performances cannot lead to safe conclusions and may therefore be abandoned.

Conclusion

The results suggest that hit rate excesses in multiple choice tests obtained by samples of participants completing ESP tests alone at home, without control, do indicate paranormal power since the majority of high scorers at home are capable of producing significant hit deviations as well under control.

One might also conclude that the test strategy applied in these tests, a *first-home-then-lab* strategy, might be recommended as a methodological rule: Experimental psi research should preferentially be conducted with selected samples of psi-gifted individuals. Psi-gifted individuals are rare. Nevertheless, they may be detected, as this study shows, almost without costs, by using appropriate tests, by letting them test themselves at home, by testing high home scorers in the lab, and by inviting successful lab performers subsequently to participate as subjects at further parapsychological research.[9] The para-community's and their critics' common lamenting about "tiny" and "elusive" experimental psi effects might eventually lose its grounds and henceforth be regarded as symptom of a methodological fallacy of the past.

Contemporary researchers should have followed, in principle, model designs of pioneers of parapsychological research who gave many examples of continuously testing psi-gifted subjects over long periods who showed considerable consistency of performance (Martin & Stribic, 1938, 1940[10] , Birge, 1942, Ryzl, 1966, review of results from star subject Pavel Stepanec, Kelly, 1972, Musso, 1973, Houtkooper & De Diana,1977, Schouten &Kelly, 1978, Sasaki & Ochi ,1982 , Hearne, 1984, Jacobs, 1985, Sijde, 1987). Replicability of performance, provided that the testing conditions do not change, has also been claimed elsewhere (example: Schmeidler, 1964).[11] Recommendations to conduct psi-tests with high-scoring participants whose performance is more reliable than that of low-scoring participants have been given early on (Schmeidler, 1948, Murphy, 1969) [12]. The complementary advice to forego unselected participants altogether if "fact-finding repeatability" is needed is not novel either (Whiteman, 1972). [13] It seems as if insights of past generations have been unduly sacrificed in favor of fashionable developments, concurrent detrimental effects on experimental psi manifestation went unnoticed.

References

Anderman, E. M., Griesinger, T., & Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology 90*, 84-93.

Bierman, D. J. & Gerding, J. L. F. (1991). Towards a reduction of experimenter control in the study of special subjects. *The Parapsychological Association 35th Annual Convention. Proceedings of Presented Papers.*

[9] This rule implies that high scorers in the balls or pearls drawing tests exhibit generalized psi ability that should manifest itself as telepathy, precognition, PK etc. under variable conditions. The predictive power (validity) of ball drawing test results has been investigated with a smaller number of high scorers and has been sufficiently confirmed (to be published elsewhere).

[10] J. B. Rhine 1977) was aware of the importance of testing psi-gifted subjects, but he apparently thought that "special" subjects would show up by luck: "Among other advantages, the use of a special subject considerably reduces the *uncertainty and loss of time required by the search for a selection of subjects from the general population* (italics by SE)....Perhaps the main lesson to be learned ... is how best to obtain and prepare subjects from the general population, and, on the other hand, how best to find the good subjects and keep them good." (Rhine, 1977, p. 44). But Rhine did not actually teach us such lesson.

[11] "*The author concludes that there are detailed, meaningful, and self-consistent patterns within ESP data which can be obscured if all of a subject's responses are pooled*" (Schmeidler, 1964, abstract). "*In regard to future work in parapsychology, we must move flexibly to new problems that are related to three main issues: First, finding the most favorable conditions under which gifted subjects can do their best work ...*" (Murphy, 1969).

[12] "...it is suggested that any experiments with gifted individuals could probably be completed with greater clarity and efficiency than those with subjects chosen at random. It may even be that sensitive subjects could disclose energy relations or physiological correlates of psi that otherwise would be imperceptible because our instruments were not sufficiently delicate or our research accurately enough aimed to detect such faint effects without them." (Schmeidler, 1948, abstract).

[13] ...Unselected subjects should not be used in experiments aimed at fact-finding repeatability..." (Whiteman, 1972, abstract). Repeatable experimental conditions are indispensible for replicating experimental outcomes (Hövelmann, 1984). Recruiting participants with sufficient psi-ability is one of the foremost – and repeatable – experimental conditions enhancing replicability of paranormal effects.

Birge, W. R., & Rhine, J. B. (1942). Unusual types of persons tested for ESP. I. A professional medium. *Journal of Parapsychology, 6*, 85-94.

Bleksley.A.E.H. (1984). Summary of an experiment with a gifted subject on long-distance ESP during sleep. *Parapsychological Journal of South Africa, 5,* 85- 99.

Brown, G. S. (1953). Probability and psychical research. *Nature, 172,* pp.154-156.

Brugger, P. & Taylor, K. I. (2003), 'ESP. Extrasensory perception or effect of subjective probability?', *Journal of Consciousness Studies,10*, pp. 6-7. Editors: Alcock, J. E., Burns, J. E., & Freeman, A., Title of issue: Psi wars. Exeter, U.K.: Imprint Academic.

Ertel, S. (submitted). Are ESP test results stochastic artifacts? Brugger & Taylor's claims under scrutiny.

Ertel, S. (in print). The ball drawing test. Psi from untrodden ground. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the 21st Century: The Future of Psychical Research.*

Feller, W. K. (1940). 'Statistical aspects of ESP'. *Journal of Parapsychology,* 4, pp. 271-298.

Gerding, J. L. F. & Bierman, D. J. (1997) Theory-based control in an unattended forced-choice psi study with a special subject. The Electronic Journal of Anomalous Phenomena, 97.1.

Milton, J. (1996). Establishing methodological guidelines for ESP studies: A questionnaire survey of experimenters' and critics' consensus. *Journal of Parapsychology* 60, 289-334.

Goodfellow, L.D. (1938), A psychological interpretation of the results of the Zenith radio experiments in telepathy. *Journal of Experimental Psychology, 23,* 601-632 (reprinted in *Journal of Experimental Psychology: General*, *121*, 1992, pp. 130-144).

Hansen, G. P. (1990). Deception by subjects in psi research. *The Journal of the American Society for Psychical Research, 84,* 25-80.

Hearne, K. M. T. (1984). A survey of reported premonitions and of those who have them. *Journal of the Society for Psychical Research, 52*, 261-270.

Houtkooper, J. M., & De Diana, I. P. F. (1977). Relationships between PK and physiological variables in a gifted subject. *European Journal of Parapsychology, 2*, 68-79.

Jacobs, J. C. (1985). Psi-guided awakening from sleep. 1: The original experiments of W. Van Vuurde. *Journal of the Society for Psychical Research, 53*, 159-168.

Kelly, E. F., & Kanthamani, B. K. (1972). A subject's efforts toward voluntary control. *Journal of Parapsychology, 36*, 185-197.

Martin, D. R., & Stribic, F. P. (1938). Studies in extra-sensory perception: II. An analysis of a second series of 25,000 trials. *Journal of Parapsychology, 2*, 287-295.

Martin, D. R., & Stribic, F. P. (1940). Studies in extra-sensory perception: III. A review of all University of Colorado experiments. *Journal of Parapsychology, 4*, l59-248.

Musso, J. R., & Granero, M. (1973). An ESP drawing experiment with a high-scoring subject. *Journal of Parapsychology, 37*, 13-36.

Pöppel, E. (1967). Signifikanz-Artefakte in der experimentellen Parapsychologie. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, 10, pp.63-72.

Reinsel, R. (1999) Advice for newcomers to parapsychology. Preliminary notes on experimental procedure for GESP studies. http://www.parapsych.org/PDF/GESP_research_guidelines.pdf

Rhine, J. B. (1977) History of experimental studies. In B. B. Wolman (Ed.) *Handbook of parapsychology*. Pp 25-55. New York: Van Nostrand .

Schmeidler, G. R. (1948). Research projects in parapsychology. *Journal of Parapsychology, 12*, 107-113.

Schmeidler, G. (1964). An experiment on precognitive clairvoyance. Part II. The reliability of the scores. *Journal of Parapsychology, 28*, 15-27.

Sijde, P. C. (1987). Pavel Stepanek, a gifted subject. *Tijdschrift voor Parapsychologie, 55*, 23.

Ryzl, M. (1966). A model of parapsychological communication. *Journal of Parapsychology, 30*, 18-30.

Sasaki, K. S., & Ochi, Y. (1982). Observation of the deformation of pure aluminum plates by psychokinesis. *Journal of the Psi Science Institute of Japan, 7*, 7-13. 7 illus; 13 ref.: Journal of the Psi Science Institute of Japan.

Schouten, S. A., & Kelly, E. F. (1978). On the experiment of Brugmans, Heymans, and Weinberg. *European Journal of Parapsychology, 2*, 247-290.

Timm, U. (1968) Das „Round-Table-Experiment" – Ein neuer Versuchsplan für ASW-Gruppenversuche. . *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie.* 11, 34-43.

Whiteman, J. H. M. (1972). On the concept of repeatability in scientific experimentation. *Journal of the American Society for Psychical Research, 66*, 227-229.

Address for correspondence: Suitbert Ertel,
Georg-Elias-Müller-Institut für Psychologie,
Gossler Strasse 14, D-37073 Göttingen, Germany.
Email: sertel@uni-goettingen.de

DOES PSI EXIST AND CAN WE PROVE IT? BELIEF AND DISBELIEF IN PSYCHOKINESIS RESEARCH

Eckhard Etzold
Gesellschaft für Anomalistik, Heidelberg

ABSTRACT

Psychokinesis research is encountering difficulties in replicating its findings. Centuries before psychokinesis phenomena such as levitation or moving objects were well known. Today anomalous phenomena seem to be reduced to rare and weak effects in stochastic processes. While experimental and analysis methods became more and more professional in the last decades, researchers complain about a loss of effect size and evidence. The "Model of Pragmatic Information" (MPI) by Walter von Lucadou predicts a change or decline of effect size. According to MPI this is necessary because otherwise a paranormal experiment could be used for information transfer which might be an intervention paradox. In elaborating further theoretic implications and consequences of the MPI we show that we finally reach a point where the outcome of a given PK experiment can not be distinguished from random results. Therefore, we should abandon proof-oriented research. All we can find are different degrees of evidence.

Another interpretation is that increasing skepticism for itself might be the reason for the erosion of evidence. The author's earlier analysis of Fourmilab Retro-PK data with respect to lunar phase yielded a z-value of 3.24 for the full moon interval. A replication yielded a z-value of -2.49 for the same interval. Some evidence is given that this effect overturn (change of sign) depends not only on the predictions of the MPI but on believing and disbelieving in paranormal phenomena. Other parapsychologists noticed that their experimental results often corresponded to their own belief or disbelief in paranormal phenomena. This seems to be more than just a coincidence or the result of single experiences, it might be part of the nature of psychokinesis phenomena itself. Such experiences are the result of interactions between one's mind with the physical world, analysed by an experimenter. In the classical scientific view the experimenter is a neutral observer of the experiment without any influence on the result. This view has to be corrected. The experimenter has also expectations, fears and hopes in his mind which could influence the outcome of a PK experiment. He also belongs, with respect to MPI, to the psi agents of a PK experiment. More than in any other scientific discipline the result of an experiment depends on the experimenters belief or disbelief in paranormal phenomena. Both, belief and disbelief, are self-referential, they act as self-fulfilling prophecies and tend to create their own evidence which confirms the expectations of the paranormal-believing experimenter like the skeptic experimenter. Beside the parameter of Lucadou's new experimental paradigm it is necessary to document the experimenter's belief in paranormal phenomena and to evaluate its effect and its outcome. The best conditions for growing evidence might be to use test subjects and experimenters who do not doubt in the existence of psi. The demand of skeptics to ban parapsychology from the realm of science have to be rejected. It is a science with its own special research conditions.

INTRODUCTION

The scientific status and position of parapsychology in the sphere of science has been a bone of contention from the very beginning (Bauer, 1985; Palmer, 1990; Alcock, 2003; Parker, 2003; Parker & Brusewitz, 2003; Irwin, 1989; Hoebens, 1982). Exponents of the skeptics' organisation GWUP (the German CSICOP) challenge the 'scientific nature' of parapsychology and seek to ban it entirely from the sphere of science if it fails to provide proofs for the existence of psi. In this context, much is made of the fact that parapsychology has yet to succeed in identical replicating a single anomalous effect under laboratory conditions (Hüsgen & Kamphuis, 2000). Beside the lack of replication we find (as a consequence of it) a decline of evidence and effect size of paranormal phenomena. Is this a consequence of increased skepticism in the last centuries?

At the beginning of the 17th century there was no academic parapsychological research. Miracles and paranormal events were generally accepted and widely evident in the society. Skepticism was just beginning to be a part of scientific work. In this time, the Italian monk St. Joseph of Cupertino, provoked the displeasure of the Holy Inquisition through the numerous cases of him levitating during the elevation of the host which could not be explained scientifically:

> There are many skeptical witnesses of the numerous levitations of Joseph of Cupertino who did not trust these phenomena and had enough scientific knowledge to justify their doubts. Yet it was precisely before such skeptical witnesses that Joseph of Cupertino levitated to such amazing heights, virtually every time that mass was celebrated. The levitation occurred in him so frequently and led to such a disturbance of the service that he had to be tied down with lead boots; yet this was to no avail and he rose together with the lead boots. Sometimes he levitated to the ceiling of the church and it was only with the greatest of effort that he could be brought down to earth again from the highest ledge, which he held on to after his awaking from ecstasy. On several occasions, an acolyte tried to hold him down but was himself carried upward together with him. (Benz, 1969, p. 218).

Makro-PK phenomena like levitation were evident in the 17th century. There was little doubt about it, and even some skeptical witnesses were convinced by the experience of paranormal phenomena.

On the wake of the 20th century, paranormal effects became slowly an area of research. Reports of poltergeist phenomena and makro-PK events were widely discussed but never got an academic status or scientific recognition. Nevertheless, mediumistic phenomena were fascinating the academic world and attracted respected scientists like the radio pioneer Oliver Lodge. The famous German author and Nobel laureate Thomas Mann only hoped to see "once again, with my own eyes, the handkerchief ascending into the red light" (Mann, 1983, p. 255), and in 1922 mediumistic-talented test persons could move macroscopic objects many feet by psychokinetic influence (Bender, 1966, p. 496). As J. B. Rhine introduced scientific methods in parapsychology to evaluate makro-PK effects with psychokinetic talented test persons trying to influence dice tossing. Later Helmut Schmidt introduced electronic devices and random event generators as targets for micro-psychokinetic influence.

Today skeptic scientists supervise every paranormal experiment design to protect it against fraud or misinterpreted natural explanations. At the same time in parapsychological research anomalous PK phenomena become rare and weak, and have shrinked to minor statistical mean value deviations and micro-PK effects in large databases containing abstract columns of numbers. It is required to run tens of thousands, even hundreds of thousands of experiments before any significance becomes apparent. The days of flying monks and PK-moved objects are over. Why did the effects loose their impressive strength? Were they all the result of fraud?

Skeptic scientists argue that with improved methods of analysis and evaluation many errors and artefacts were excluded which seem to be the true source of claims of the paranormal. When the highest standard of analysis is reached, no paranormal phenomena will remain. But this is only one interpretation. It is the aim of this presentation to introduce another interpretation: increasing skepticism for itself might be the reason for the erosion of evidence. This depends on the nature of paranormal phenomena itself.

THE LACK OF SUCCESSFUL REPLICATIONS

In 1997 the Princeton Engineering Anomalies Research group (PEAR) published its evaluation of a twelve-year series of micro-psychokinesis tests with random number generators (RNGs) which came to a (statistically) impressive conclusion:

> The overall scale of the anomalous mean shifts are of the order of 10^{-4} bits per bit processed which, over the full composite database, compounds to a statistical deviation of more than 7s ($p = 3.5 \times 10^{-13}$) (Jahn et al. 1997, p. 363).

The effect size of one bit in every 10,000 which could be changed by the test subject in the intended direction appeared to be reliable and leading to the expectation that psychokinesis really exists as an anomalous, replicable phenomenon. A similar conclusion was drawn by Dean Radin:

> After sixty years of experiments using tossed dice and their modern progeny, electronic RNGs, researchers have produced persuasive, consistent, replicated evidence that mental interaction is associated with the behavior of these physical systems (Radin, 1997, p. 144).

This leads to the expectation that PK effects could be easily reproduced with a large number of tries and test subjects. In 1996 the Fourmilab Retro-Psychokinesis Project was founded, an internet based psychokinesis experiment which opened the possibility for interested test persons all around the world to take part in the PK experiment. But the still ongoing Fourmilab Retro-Psychokinesis Project failed in generating significant mean shift deviations in the overall summary. 162,687 registered experiments since January 11th., 1997 were counted until June 10th, 2004, each of it with 1,024 bits and a total sum of 166,591,488 "tries" performed by 15,686 test subjects. According to PEAR's expectation of one bit per 10^4 bits changing in the intended direction meanwhile the total z value has to be appr. 2.58 standard deviations. Instead of it the actual overall z reached a non significant level of -0.4377 standard deviations.

Also in 1996, the collaborative program of anomalous Mind-Machine Interactions (MMI) under the leadership of the PEAR group was established. The laboratories of the Freiburg Anomalous Mind-Machine Interaction (FAMMI) and the Giessen Anomalies Research Program (GARP) took part in it. Their common goal was to replicate the successful PK results of the PEAR PK experiments. What, then, could be more disappointing than to discover in the years that followed that the large-scale replication test performed by the MMI consortium was neither able to confirm the effect size that had previously been established nor to attain the level of significance which was to be expected on the basis of the tests run previously (Jahn et al. 2000).

This disappointment certainly fed the (skeptical) suspicion that anomalous psychokinesis effects do not exist (Alcock, 2003). However, according to Walter von Lucadou there would have been in fact no real reason to be disappointed, if the MMI consortium had applied his model of pragmatic information to the replication and the formulation of the effect size expectation. His model predicts that decline effects must arise in future replications (Lucadou, 2001). Is the model of pragmatic information convincing enough to reject any skeptic objection?

THE MODEL OF PRAGMATIC INFORMATION

The model of pragmatic information (MPI) is a theoretical approach predicting such declining effects in psychokinesis experiments. It is not yet a complete and finalised theory, merely a model which seeks to describe with analogies the conditions in which an anomalous effect might be expected.

In MPI anomalous or psi effects are not supernatural but meaningful correlations between the test person (psi agent) and the target system (RNG). While interacting, the psi agent (or test subject) and the RNG become a closed system with self-referential dependencies, an 'organisational closure' (Varela, 1985). This is irrelevant to any temporal or spatial distances, it is a non-local analogy to non-local effects in quantum mechanics (Lucadou, 1992). Its boundaries are defined by the ratio of internal and external pragmatic information in the interaction of its constituent parts. (Lucadou, 2001) The non-local correlations of MPI are – in worst case – only a weak violation of the laws of nature as known today because the underlying mechanism of the correlation is unknown. However, the situation becomes more critical when such non-local correlations are supposed to be used for long-distance transfer of information or signals. The possibility of intervention paradoxes prohibits such an information transfer: it would be a serious violation of natural laws. (If I know what will happen in the future I can act in the present in such a way that I can prevent unpleasant future events occurring.) Therefore Lucadou recommend: "Do not treat psi as a signal!" (Lucadou, 2001, p. 10)

Pragmatic information is "a measure for the meaning of the information". It manifests itself in "its effect on the system", but it has no informative content (unlike a newspaper or a newscast on the radio).

Pragmatic information (I) which a system produces, is in itself the product of further factors which exclude the possibility of using pragmatic information for signal transfer: An event with the character of novelty happens unexpected and suddenly, it cannot be the basis of signal transfer. An event which acts with autonomy cannot use for signal transfer too. These factors of pragmatic information exist in opposites: Novelty (E) vs. Confirmation (B). (Weizsäcker, 1974) Lucadou added later Autonomy (A) vs. Reliability (R).

(Lucadou, 1997) . The portion of pragmatic information grows in line with an increase in the portion of autonomy and/or novelty. The system itself contains something that resembles a "memory" in which the system states of the past are "stored". While the factors B (Confirmation) and "R" (Reliability) rise, the product I (the produced pragmatic information) falls. These factors are responsible for the decline effects observed in the replication experiments, because the novelty declines when repeated tests are run to reproduce such effects. At the same time, the autonomy is limited, since one possible test result is already available as a result of the pilot experiment. In order to enable a repetition of a high degree of novelty, the effect must emerge either elsewhere in a replication where it is expected, or it must change its effect size or direction. MPI provides the possibility of conceptual replications with high degrees of novelty and autonomy. Identical replications have to fail: if they are successful they could be used for signal transfer which would violate the excluding of intervention paradoxies.

After all, with such a model the results of random experiments can be described. Yet, how do we know if a single psi effect claimed is a (still) unexplainable anomaly and not simply a variety of coincidence? Can meaningful research activities be conducted at all under such conditions?

In 2000 the author made a conceptional replication of Radin and Rebman's claimed full moon effect in casino payout rates (Radin & Rebman 1998) with Retro-PK experimental data from Fourmilab (Watkins, Moore & Walker, 1996), expecting that the Fourmilab Retro-PK data would demonstrate the same full moon effect which was claimed by Radin and Rebman. The time serial analysis of the experimental data with respect to lunar phase was published in the year 2000 (Etzold, 2002). For the full moon interval (+/- 1 day) A significant *z*-value of 3.24 for the first 53,082 Fourmilab Retro-PK experiment data seems to confirm Radin and Rebman's claims of a peak effect in the period of one day before and one day after full moon. Was it an anomaly or just a coincidence? After the publication in 2000 I made a replication of my first analysis with the next 47,192 experimental data which were accumulated in the Fourmilab Retro-PK data base until August 2001. This time I was doubtful about the outcome of the analysis. I could not believe that the observed lunar effect was persistent enough for replication (MPI for example forecasts a decline effect for the new evaluation). Now I got a (negative) *z*-value of -2.49 for the specified full moon time period, and I reported that this replication failed (Etzold, 2002).

Referring to my results (Etzold, 2002), von Lucadou wrote (2002, p. 83):

> The MPI ... does not state that, if the experiment were to be repeated, the effect that had been established earlier would simply disappear, since it was merely a random fluctuation. Under MPI, it either disappears *slowly*, something which one would not normally expect to occur with a random fluctuation, or it *overturns* (as was the case in the Etzold study), or it appears in *other channels*, as occurred during the large-scale MMI replication experiment (Jahn et al. 2000).

In a somewhat schematised form, three possibilities therefore emerge under MPI for an anomalous effect during replication:

a) Slow reduction (Decline)
b) Overturn, change of signs
c) Emergence in 'other channels' (Displacement)

Do these truly represent all of the possibilities, or are there more? Lucadou certainly has excluded the possibility of only a purely random result presenting itself during replication. As far as the three possibilities are concerned, a) and b) would appear to be reasonable to the extent that they are found in the observation direction or at the other end of the scale. Yet, possibility c) appears to be highly problematic. How do I know in which 'channel' the effect will re-appear? What happens if I am unable to find the channel because I do not possess the methods and measuring techniques for this channel?

These three possibilities therefore are not a real help if I am unable to say immediately after completing the replication experiment and prior to evaluating the data whether or not, under the circumstances, I can expect the outcome to fall into category a), b) or c). Without further definition, the three possibilities put forward by Lucadou can be applied to the expected effect of any given RNG experiment in replications. A lack of evidence always remains. Lucadou himself admits in general (2001):

> To my conviction, parapsychology has ... not yet succeeded in establishing indisputable scientific evidence that psi exists. (p. 7)

In a discussion with Volker Guiard (Lucadou, 2003), Lucadou points to the two fundamental theorems of parapsychology which I would like to reiterate at this juncture (Lucadou, 1997, p. 162):

1) Psi phenomena are non-local correlations in psychophysical systems that are induced through pragmatic information which is generated by the (organisationally closed) system.

2) Each attempt at using non-local correlations for the purpose of signal transmission causes these to disappear, or converts them in an unpredictable manner.

In connection with the second fundamental theorem and its implied avoidance of intervention paradoxes, Lucadou (2003) also writes

> that psi must be conditioned in such a manner that no reliable signal transmission can result. This would suggest that, during a psi experiment, each statistical deviation that is measured and which can be interpreted as psi or an anomaly may not exceed a certain parameter (p. 139).

A *signal transmission* would mean: a clear and identificable signal which is more than pragmatic information without any uncertainty.

WHY THE EXISTENCE OF A PSI-ANOMALY CANNOT BE PROVEN

In empirical science, inductive evidence is taken to confirm hypotheses which are derived from experience, observations and experiments. In this context, the term 'inductive' merely stands for a probable causal link between a hypothesis and the findings of an experiment or observation. The amount of truth in a hypothesis becomes all the more probable, the more frequently it can be repeated. Evidence relies on information which can be obtained from evaluating the experimental data.

For parapsychology this process of obtaining evidence depends fundamentally on MPI:

> Because the MPI is a general system-theoretical description of interacting (self-referential) systems, it can also be applied to the system that creates scientific evidence. (Lucadou, 2001, p. 10)

The information contained in the claims of evidence can, for example, be summarised in one sentence: 'anomalous psi-phenomena exist'. This is more than just external pragmatic information. It is a concrete piece of information content. This means that the correlation must be so convincing that it unmistakably 'conveys' such information and consequently assumes the character of a signal. This approach, however, violates the second fundamental theorem of parapsychology since, after all, the intention of this 'horizontal signal transmission' is to convey the information that 'the anomalous psi-phenomena exist'. The consequence of this is that the non-local correlation disappears or is modified in an unpredictable manner. In concrete terms, this means, that as soon as the experiment is repeated for the purpose of proving the anomaly, the results of the experiment will vary in the frames of the null hypothesis.

For skeptics the *conditio sine qua non* for evidence is replication, and psi research can never achieve the status of science because these phenomena cannot be replicated. Hergovich (2001) summarises the skeptical position:

> To date, no convincing experiment has been put forward that proves the existence of psi-phenomena. Not because the methods required by psychology could not be met or because the effect sizes were perhaps too weak ..., but because the effects are not reliable enough. (p. 122)

Under MPI, however, the effects cannot be 'reliable enough'. The situation becomes even more complicated. In such an experiment, in every fragment of evidence the whole questionableness of our conventional worldview is present. With such a burden of information, the replications possibly not produce anything other than random fluctuations, according to the second fundamental parapsychological theorem.

Under scientific research conditions, psi comes therefore in fact across as a troll, a ghost that only manifests itself when there is no scientific conclusiveness. "The more confident one is of having 'bagged' the psi effect, the lower the chances are that it can be replicated in a future experiment" (Lucadou, 1997, p. 187). However, on this basis, it is no longer possible to prove psi-anomalies through further replication with the help of scientific laboratory research, and any attempt will lead to further disappointment. What ways out are there?

BELIEVE AND DISBELIEF IN PARAPSYCHOLOGICAL RESEARCH

Anomalies in the sense of psi effects are evidently phenomena with the property that they cannot be proven using conventional scientific methods. With respect to psi anomalies we have to look for evidence but not for proofs because proof-testing methods will destroy any possibility of finding evidence. Lucadou (2001, p. 13) has therefore proposed a new experimental paradigm that has been derived from MPI and which modifies the exterior test procedures and their evaluations with a view to attaining better findings. These include among others: no accumulation of evidence; short test runs; triple blindness; conceptual, i.e. no identical replications. However, his fundamental requirement alone, that of not treating psi as a signal, raises doubts as to whether this new paradigm can produce better results. As long as this new paradigm is also accompanied by a level of interest in producing scientific proofs, any potential anomalous effect may be bound to collapse, no matter how much autonomy and novelty the experiment is subjected to.

If the core statements of MPI and the conclusions that we have drawn here are correct, we are dealing with a class of phenomena that, per se, cannot not be proven by conventional means. All we can get are different degrees of evidence.

Dealing with psi anomalies scientifically therefore requires that I already believe in the existence of these anomalies if I want to obtain positive and significant results. This credo is not to be interpreted as any form of intellectual shortcoming like Hergovich (2001, p. 171) claims, but rather as an opportunity. By doing this, the experimenter is taking off the pressure from his research activities to have to prove something that cannot be proven. This could facilitate the scientific progress in parapsychology. The traditional Cartesian doubts that are prevalent in natural science are also merely a subjective fundamental principle which is just as capable of producing its 'cognitive blind spots' as 'belief' does. Yet, in the case of parapsychology, the Cartesian doubts are counterproductive, as it has been shown at last by the failed replication tests performed by the MMI consortium (Jahn et al. 2000). Only by this way it can be checked if the claimed human-machine interaction actually exists, if the thoughts in the mind of the experimenter can generate a corresponding effect in the physical world. Those researchers who believe in the existence of anomalous phenomena will get more positive results in PK experiments with other test persons (Smith, 2003). Those who doubt this will get also the appropriate 'psychokinetic result' which seems to negate the existence of paranormal phenomena. The growing lack of positive PK results, the "erosion of evidence" (Lucadou, 2001, p. 7) might be a result of growing disbelief in the possibilities of PK which for itself could be a PK-generated result.

The first true indication of this effect emerged in the studies performed by Gertrude Schmeidler (1943) on the effect of belief or disbelief persuasions in ESP experiments. She observed that subjects who believed in an anomalous effect (the 'Sheeps') performed better than those who viewed anomalous effects with skepticism (the 'Goats'). Schmeidler's notion of separating the 'sheep' from the 'goats' was: "Do you believe it is possible that ESP can be shown under the condition of this experiment?" A meta-analysis of the 'sheep-goat ESP studies' for the years between 1947 and 1993 performed by Lawrence (1993) produced an astronomically high *z*-value of 8.17 ($p=1.33 \times 10^{-16}$) which provides high evidence for the existence of a sheep-goat effect. Edgar Wunder complements in reaction to my own reflections (Etzold, 2004):

> The meta-analysis of Lawrence already was even a successful replication, namely of the above comparable study of Palmer (1971). Palmer (1971) found in the studies published till there a sheep-goat effect of a medium effect size which Lawrence found in the studies published afterwards again in the same order of magnitude.

At first sight it seems that contrary to the predictions of the MPI successful replications are possible. But we have to state that the sheep-goats meta-analysises lie beyond the scope of the MPI. Belief is a category which amplifies the character of the closed system, the organisational closure, which is the basic condition for the MPI. Disbelief of the psi agents prevent the development of an organisational closure. In this sense the high evidence for the sheep-goats effect might also be an evidence for the MPI. Belief and disbelief are the basic categories which allow or prevent that MPI becomes effective. But successful results are limited by becoming character of a proof. With other words: MPI with belief as a basic condition might work as long as one does not "treat psi as a signal". We will find (more or less) growing evidence but no undoubtful proofs of the paranormal.

The same statement appears to be applicable to scientists who perform research in this field. In the classical scientific view (and in Lucadou's MPI) the experimenter himself is neutral and objective regarding any results of the experiment. Only the external pragmatic information which is generated by the closed system is of importance. But with increasing interest in the outcome of the experiment, the experimenter himself can become part of the organisational closure too and interact with the target system. Regarding my own studies (Etzold, 2000, 2002) I wonder who were the psi agents in the case of the observed lunar effect in the Fourmilab Retro-PK data, the approximately 8,000 test subjects who didn't and still don't know that they were tested for lunar effects, or I, the experimenter, who believed / disbelieved in lunar effects? Believing in psi seems to improve the results of PK experiments. Parker noted:

> Recently, Matthew Smith and Michael Gordon investigated the psychology of the 50 named 'psi-conducive and psi-inhibitory experimenters' and found by multiple regression of self-report questionnaires that higher psi-conduciveness scores were associated with belief in one's own PK [psychokinetic] ability (Parker, 2003, p. 128)
> [and added:]
> Some empirical support is found in a review by Brian Millar … who concluded that considering psi ability is rare, psi-conducive experimenters were themselves to be found over-represented as psi-conducive subjects!

Smith, while discussing different kind of experimenter effects based on social-interactional explanations (Smith, 2003), has collected some successful studies of parapsychological experiments which might confirm this statement and supposed:

> If psi is real, then it is plausible, indeed likely, that the experimental participants are not the only source of psi in a successful parapsychology experiment. The experimenter may also exert a psi influence over the data. Given that apparently 'psi-conducive' experimenters typically tend to believe that psi exists, and are highly motivated to obtain findings in support of psi (often more so than their research participants) then one might argue that the experimenters are potentially a more significant source of psi than the participants. (Smith, 2003, p. 79)

Others before him have suggested the same experimenter-influence and noticed some anecdotic material:

> For example, when Blackmore, a devoted parapsychologist for many years, found herself increasingly skeptical about *Psi* as a consequence of her inability to produce experimental evidence for it, she noted that 'many parapsychologists suggested that the reason I didn't get results was quite simple – *me*. Perhaps I didn't sufficiently believe in the possibility of *Psi*' (Alcock, 1987, p. 561).

This is possible. Lucadou wrote in view of the MPI: "… the model also includes the reverse action of pragmatic information from outside to inside." (Lucadou, 2001, p 11), and Smith (2003) commented in view of experimenter effects:

> From a methodological perspective, whatever the purported mechanism(s) of this effect of the experimenter upon the data, it does raise potential problems for skeptical researchers who wish to attempt to replicate psi experiments. This is because it suggests that such researchers, especially if they act as the experimenter who comes into contact with research participants, are less likely to obtain positive findings even if the psi effect is real. (p. 82)

This material gives some evidence for the claims that a causal link exists between the decline of effect size and erosion of evidence with increasing of scientific criticism and skepticism. If this is true one skeptic experimenter or even other persons like checkers or observers (White, 1976a) could dominate the effect size of the whole experiment.

Alcock (2003) told an example for this case in which his friend Jeffers was involved, but without noticing that he himself could be the reason for obtaining negative results. "Jeffers stands in lonely company as one of the very few *neutral scientists* who have empirically investigated the existence of psi phenomena." (Alcock, 2003, p. 36) Jeffers tried a conceptional replication of the PEAR RNG-PK experiments, not using RNGs but interference of light as target for anomalous influence. Alcock himself, whose position is radical skeptic, was involved in this experiment:

> Jeffers came to me at least a tad defiantly, requesting that I review his experimental design and offer any suggestions and criticisms before he began his research. He stressed that I should not after the fact, were he to obtain data supporting the parapsychological interpretation, then argue that the experiment was not to be taken seriously because it had fallen methodologically short in some fashion. Thus began our relationship, which was to grow into the very positive one that it is today. (Alcock, 2003, p. 36-37).

In the term of the MPI Alcock himself became part of the organisational closure, in this case as a doubtful experimenter who wished to find the confirmation for his disbelief in Jeffers' experimental result: "As Jeffers reports in his paper, his research findings give no support to the Psi hypothesis." (Alcock, 2003, p. 37) The possibility that Alcock himself produced via the experimenter's psi faculties the negative result of Jeffers' research was not discussed in his paper, but cannot be ruled out if we apply the MPI for the whole system which consists of Jeffers, his experiment target and also Alcock as critical designer and reviewer of the experiment. Alcock, who believes in the null hypothesis and asks to give the null hypothesis a chance will find nothing else than evidence for the null hypothesis. If psi exists, and I believe it, psi will also acting in the skeptics attempt to obtain evidence for the non-existence of psi.

CONSEQUENCES

In science we have "two schools of research on belief in the paranormal" (Lawrence, 1993, p. 83), represented by scientists and investigators who differ fundamentally in their approach: "'Parapsychologists' who believe in the possible existence of anomalies as well as the 'skeptics' who reject the idea that anomalies or paranormal phenomena could exist (Hergovich, 2001, p. 119). Every school has their own lists of studies which provide evidence for the correctness of their own belief or disbelief. These two schools have been around since the inception of scientific parapsychology, and they are testimony to the fact that the scientific status of parapsychology was undefined in the beginning.

The conclusion drawn by the parapsychologists that predicated anomalies (or psi effects) cannot be proven in sense of a skeptical proof, might alleviate the tension in the relationship. For the 'skeptics', this would mean making a concession of not demanding from the parapsychologists what they themselves (and other scientific disciplines) can not produce. For the parapsychologists, it would mean relief in that they would no longer need to have to 'prove' anything to 'the others'. Instead of having to invalidate their own findings in a proof-orientated world of research, they have now found space to run process-orientated research. This means that they no longer seek to prove whether or not an anomalous effect actually exists but involve themselves in an anomalous phenomenon and initially describe what experience on this effect is being gained in the field of scientific research.

Lawrence (1993) claimed:

> What is needed is a good, reliable, accurately validated measure of general belief in the paranormal (...). Questions should most certainly include the Schmeidler question seen to be joint most successful measure of belief in terms of getting results. (p. 83)

Together with Lucadou's requirements of no accumulation, short runs, conceptual replications (Lucadou, 2001, p. 13) it is necessary to add the requirement of believing in successful PK experiments to the MPI: "It is obvious that the role of the experimenter (conceiving this term in its broadest sense) must be taken into account in designing the results of parapsychological experiments" (White, 1976b). And Parker (2003) added: "High-scoring subjects and successful experimenters are to be found and a technology is available." (p. 132) Test subjects like experimenters should be tested before the beginning of an experiment, using a variant of Schmeidlers question: "Do you believe it is possible that PK can be shown under the condition of this experiment?" For doing successful parapsychological laboratory work it seems necessary and consequent to document the belief or disbelief of the experimenter for further evaluations.

More than in any other scientific discipline the researcher and the experimenter themselves are part of the experiment they observe and analyse. Their expectations, hopes, fears, belief and disbelief are self-referential, they act as self-fulfilling prophecies (Watzlawick, 1985). They may influence the outcome of RNG experiments in the same manner as the attempts of the test persons to influence the random processes of the RNG. The experimenter, regardless of his beliefs, has probably the highest interest of all in the outcome of the experiment. Therefore he might be the most powerful psi acting agent – possibly against his own will.

The existence of anomalies or psi effects cannot be proven as we have seen. Everybody is likely to find evidence for his own belief. It is equally possible to gather strong evidence for the existence of psi like it is possible to gather strong evidence for the null hypothesis. The one is true, and the opposite is true, too. Evidence in this case means only that belief or disbelief create their own corresponding results in the real world (Etzold, 1992). The answer for the question "Does Psi Exist?" (Parker, 2003) is undecided and has to be undecided as long as we have found no convincing evidence which might even satisfy skeptic doubts. Eberhard Bauer (1991, p. 138) states that in spite of all the skeptic doubts, parapsychology still belongs in the realm of science. For scientific acceptance now it is more important to say under which conditions the existence or inexistence of psi is falsifiable. Very general, the thesis, psi does not exist, is falsifiable if every human experience can be explained in conventional scientific terms. The thesis, psi does exist, is falsifiable if anomalous human experiences will be found which cannot be explained in conventional scientific terms. Bauer qualifies this by writing that parapsychology "does not seek to prove psi but instead wants to find explanations for a certain type of human experiences for which temporary was used the neutral theoretical term psi". (Bauer, 1991, p. 142). Parapsychology has to be considered a scientific discipline as long as human beings have experiences which can't be explained with the help of conventional scientific knowledge. However, this discipline has research approaches different from any other scientific branch. Against skeptic claims that no paranormal effects were ever replicated, we have to state that replications are possible. Parker and Brusewitz have given a list of successful research reports. The summarised results of parapsychological experimentation are indicative of an anomalous process of information transfer (Parker & Brusewitz 2003). Evaluating the state of belief/disbelief of the experimenters in connection with the experimental results might be another way for finding growing evidence. However, it is highly questionable that this will convince skeptics. We don't know actually what psi is. Perhaps all our models and reflections fall too short. Will the times of flying monks and ascending handkerchiefs return in future when the battle of skeptics and parapsychologists have finished, and teach us much more about the object of our research than we have learned in the last decades?

ACKNOWLEDGEMENTS

The author wishes to thank the referees of the program committee as well as Hugh Deasy and Udo Unrau for reading and commenting the text. The support of the Gesellschaft für Anomalistik is also gratefully acknowledged.

REFERENCES

Alcock, J. (1987). Parapsychology: science of the anomalous or search for the soul? *Behavior and Brain Sciences*, 10, 553-565.

Alcock, J. (2003). Give the Null Hypothesis a Chance. Reasons to Remain Doubtful about the Existence of Psi. *Journal of Consciousness Studies*, 10, 29-50.

Bauer, E. (1985). Gesang zwischen den Stühlen – oder wie fühlt man sich als „Parapsychologe"? In: Feyerabend, P., Thomas Chr. *Grenzprobleme der Wissenschaften* (pp. 367-373). Zürich: Verlag der Fachvereine.

Bauer, E. (1991). Zwischen Devianz und Orthodoxie. Versuch einer Standortbestimmung der Parapsychologie. In: Eberlein G. *Schulwissenschaft – Parawissenschaft – Pseudowissenschaft* (pp. , 131-146). Stuttgart: Hirzel.

Bender, H. (1966). *Parapsychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Benz, E. (1969). Die Vision. Erfahrungsformen und Bilderwelt. Stuttgart: Klett.

Etzold, E. (1992). Schafft sich der Glaube seine Wirklichkeit selbst? Religiöse Phänomene in konstruktivistischer Weltsicht. *Pastoraltheologie*, 81, 429-442.

Etzold, E. (2000). Lunarperiodische und solarperiodische Einflüsse in Psychokineseversuchen. *Grenzgebiete der Wissenschaft*, 49, 149-174.(English: http://bs.cyty.com/menschen/e-etzold/archiv/science/lunarpk1.htm)

Etzold, E. (2002). Sind die Daten der Fourmilab-Experimente mit der Mondphase korreliert? Ein Replikationsversuch. *Zeitschrift für Anomalistik*, 2, 76-90.

Etzold, E. (2004). Ist die Existenz von Psi-Anomalien beweisbar? *Zeitschrift für Anomalistik*, 4, in press.

Hergovich, A. (2001). Der Glaube an Psi. Die Psychologie paranormaler Überzeugungen. Bern: Huber.

Hoebens, H. (1982). Die Legitimität des Unglaubens. In: *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, 24, 61-73.

Hüsgen, I., Kamphuis, A. (2000). Hans im Unglück. Ein Bericht von der 43. Jahrestagung der Parapsychological Association. *Skeptiker*, 14, 199-203.

Irwin, J. (1989). On Paranormal Disbelief: The Psychology of the Sceptic. In: Zollschan, G.K.,Schumaker, J.F., Walsh, G.F. *Exploring the Paranormal: Perspectives on Belief and Experience* (pp. 305-312). Channel Islands: Guernsey Press.

Jahn, R.G., et al. (2000). *Mind/Machine Interaction Consortium: PortREG Replication Experiments*. Research Report No. 1, April 2000. Freiburg.

Jahn, R. G., Dunne, B. J., Nelson, R. D., Dobyns, Y. H., Bradish, G. J. (1997). Correlations of Random Binary Sequences with Pre-Stated Operator Intention: A Review of a 12-Year Program. *Journal of Scientific Exploration*, 11, 345-367.

Lambeck, M. (1997). Können Paraphänomene durch die Quantentheorie erklärt werden? *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, 39, 103-116.

Lawrence, T. R. (1993). Gathering in the Sheep and Goats. A Meta-Analysis of Forced-Choice Sheep-Goat ESP Studies, 1947-1993. In: Schlitz, M.J. *The Parapsychological Association 36th Annual Convention. Proceedings of Presented Papers* (pp. 75-86). Toronto.

Lucadou., W. v. (1992). Makroskopische Nichtlokalität. *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie* 34, 201-216.

Lucadou., W. v. (1997). Psi-Phänomene. Neue Ergebnisse der Psychokineseforschung. Frankfurt/Main: Insel.

Lucadou., W. v. (2001). Hans in Luck: The Currency of Evidence in Parapsychology. *Journal of Parapsychology*, 65, 3-16.

Lucadou., W. v. (2002). „Mitlaufen" des MPI ist zu begrüßen. *Zeitschrift für Anomalistik*, 2, 82-84.

Lucadou., W. v. (2003). Wie verschwindet Psi? Eine Erwiderung auf Volker Guiards „Bemerkungen zum Modell der Pragmatischen Information". *Zeitschrift für Anomalistik*, 3, 138-142.

Mann, T. (1983). *Über mich selbst*. Autobiographische Schriften. Frankfurt/Main: Fischer.

Palmer, J. (1990). Haben wir Psi nachgewiesen? Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie, 32, 6-18.

Palmer, J. (1971): Scoring in ESP tests as a function of belief in ESP. Part I. The sheep-goat effect. *Journal of the American Society for Psychical Research*, 65, 373-408.

Parker, A. (2003). We Ask, Does Psi Exist? But is this the right question and do we really want an answer anyway? *Journal of Consciousness Studies*, 10, 111-134.

Parker, A., Brusewitz, G. (2003): A Compendium of the Evidence for Psi. *European Journal of Parapsychology*, 18, 33-51.

Radin, D. (1997). *The Conscious Universe*. San Francisco: Harper.

Radin, D.I., Rebman, J.M. (1998): Seeking Psi in the Casino. *Journal of the Society for Psychical Research*, 62, 193-219.

Schmeidler, G. R. (1943). Predicting good and bad scores in a clairvoyance experiment: A preliminary report. *Journal of the American Society for Psychical Research*, 37, 103-110.

Smith M. (2003). The Role of the Experimenter in Parapsychological Research. *Journal of Consciousness Studies*, 10, 69-84.

Steinkamp, F., Boller, E., Bösch, H. (2002). Experiments examining the Possibility of Human Intention interacting with Random Number Generators: A Preliminary Meta-Analysis. Preprint, *Proceedings of the 45th Annual Convention of the Parapsychological Association.* Paris 5th to 8th of Aug. 2002.

Varela, F. (1985). Der kreative Zirkel. Skizzen zur Naturgeschichte der Rückbezüglichkeit. In: Watzlawick, P. *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus* (pp. 294-309). München: Piper.

Watkins, M. R., Moore, P., Walker, J. (1996): The RetroPsychoKinesis Project. URL: http://www.fourmilab.ch/rpkp/

Watzlawick, P. (1985). Selbsterfüllende Prophezeiungen. In: Watzlawick, P. *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus* (pp. 91-110). München: Piper.

Weizsäcker, E. v. (1974). Erstmaligkeit und Bestätigung als Komponenten der pragmatischen Information. In: Weizsäcker, E. v. *Offene Systeme I. Beiträge zur Zeitstruktur von Information, Entropie und Evolution* (p. 82-113). Stuttgart: Klett.

White, R. A. (1976b). The limits of experimenter influence on psi test results: Can any be set? *Journal of the American Society for Psychical Research*, 70, 335-369.

White, R. A. (1976a). The influence of persons other than the experimenter on the subject's scores in psi experiments. *Journal of the American Society for Psychical Research*, 69,133-166.

Wunder, E. (2004). Fragwürdige Prämissen zur Konzeption der Parapsychologie. *Zeitschrift für Anomalistik*, 4, in press.

Address for correspondence: Pfr. Eckhard Etzold,
Kirchengemeinde St. Jakobi, Goslarsche Str. 31,
D-38118 Braunschweig.
E-mail: eckhard.etzold@gmx.de

THE LOCAL EVENT DETECTOR (LED) – AN EXPERIMENTAL SETUP FOR AN EXPLORATORY STUDY OF CORRELATIONS BETWEEN COLLECTIV EMOTIONAL EVENTS AND RANDOM NUMBER SEQUENCES

Johannes Hagel, Margot Tschapke

Institut für Psycho-Physik (IPP), Köln

ABSTRACT

In this contribution we describe a conceptually new unit for the generation of sequences of binary numbers. It is based on a two oscillator system in which a low frequency component with frequency f_1 samples an alternating oscillation of a larger frequency f_2. The successive results of the sampling are directly used as binary output sequence and consequently the binary numbers appear with the period of the low frequency oscillator. In addition this oscillator shows a frequency variation which is large ($\Delta f_1 = \pm 10\ \%$ from mean value) as compared to usual circuits generating electrical oscillations. This is achieved by choosing the working point of an RC circuit such as to be located in the non linear part of the transistors characteristics. This implicates that the output sequence consists of random strings of binary numbers which are irregularly interrupted by sub-sequences containing an "over amount" of binary zeros or ones. By "over amount" we define a significant local deviation of the actual distribution from an expected Bernoullian distribution of 1-s and 0-s (more than two sigma from expectation related to the initial point of the deviation).

Like in the Global Consciousness Project (GCP) of Roger Nelson (1997) we found that resonant deviations frequently occur synchronous to emotionally charged events in the spatial neighbourhood (local Köln area) of our institute. This relates as well to important local events in and around Köln as well as to more global events of which one may suppose that they are of emotional importance for large parts of the local population too. The presumably local character of the effect becomes even more visible if we are dealing for example with important sports events (like football matches in the local stadium) In contrary, equally important events for a local population in large distances (e.g. a heavy traffic accident in Geneva) never showed the described synchronous deviation. Finally we found indications that there appears a certain experimenter effect, meaning that repeatedly we registered resonant excitations at the very instant of the occurrence of emotional important events concerning ourselves.

INTRODUCTION

Following Hagel & Tschapke (2004) a sequence of binary numbers can be generated by the following method: An electrical or electronic oscillator generates a series of impulses with low frequency which are transformed into short pulses of light generated by a light emitting diode (LED). As indicated in Fig. 1 these light pulses arrive at a receiver placed in 2 meters distance from the light source. The receiver is coupled to a second independent oscillator generating an exactly symmetrical rectangular oscillation. A special circuit (sampler) decides if the incoming light impulse arrives during a HIGH or LOW level part of the rectangular oscillation. In the first case the output is the binary number "1" while in the second case a "0" is produced.

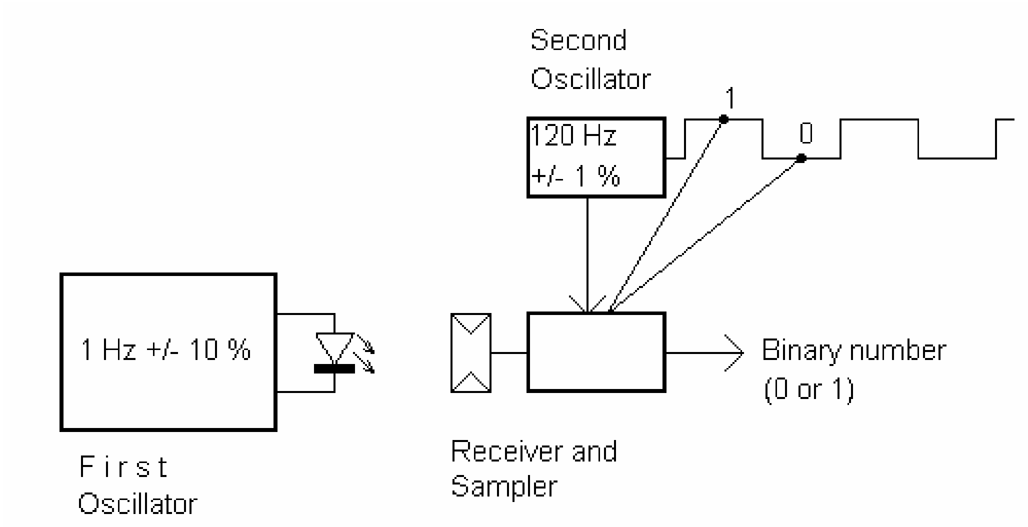


Fig. 1: Two oscillator system for the generation of binary number sequencers

The first oscillator is constructed in such a way that it has a strong frequency variation of about 10 % from its nominal value f_1 = 1 Hz. This is realized by running the unit near a cutoff condition where the non linear behavior of the underlying circuit becomes chaotic and unpredictable. The second oscillator (running at 120 Hz) produces a highly symmetric (equal lengths of high and low status – precision 50ns) rectangular oscillation. Since the two oscillators are assumed to be physically independent and due to the random frequency variation in f_1 the binary output sequence of the system becomes random in general. In general it follows a Binomial distribution as has been confirmed by simulation (Hagel, 2004). However, due to the non linear variation of f_1 it happens repeatedly that there appear transient strings of light impulses with a nearly constant frequency (due to unstable equilibrium states in chaotic systems (Lichtenberg, 1983)). If in addition this nearly constant frequency value is in resonance with the second oscillator in the sense that a mathematical relation holds of the form

$$N f_1 = f_2 \quad ; \quad N \text{ .... integer number}$$

then there appears a departure from randomness in the sense that the cumulative difference between "1" and "0" events escapes from what would be the statistical expectation in a really random system. Such resonant deviations by themselves are *NOT to be seen as "anomalous"*; they occur with a certain probability being inversely proportional to the ratio f_1/f_2 and to the amount of frequency variation Δf_1 of the first oscillator. In the GCP of Roger Nelson (1997) the interpretation is that there exist correlations between random number generators and emotional states which let the RNGs deviate from expectation. In our LED experiment these deviations are considered natural but we assume that there exists a correlation between their occurrences in time and the emotional states. In spite of this difference in interpretation the method of data analysis in LED is equivalent to the one of the GCP meaning that we identify the occurrences of deviations by excursions of the cumulative differences from Bernoullian expectation. Due to the complexity of the frequency oscillations of the first oscillator it is practically impossible to perform a strict statistical analysis of the described resonance crossings and therefore to determine strictly the average number of deviations per time unit. To overcome this drawback we performed a computer simulation of the non linear effects in the first oscillator using as input the measured non linear characteristics of the electronic circuit. The detailed results are presently prepared for publication (Hagel, 2004). The main outcome of the simulations is that for the parameters given in the subsequent section we expect a deviation from an unperturbed Bernoullian distribution of more than Z = 2 every 48 hours in average. This agrees well with our present observations. The binary numbers are stored continuously and the cumulative difference (CD) between the number of "0" and "1" events is computed and stored in addition. The CD is then investigated and checked for departures from expectation linked to the random strings. We believe that the LED set up can be considerably more sensible to correlation effects (if there are any) than a single classical RNG in the GCP: Consider a classical RNG (e.g. based on radioactive decay) supposedly correlating with an external emotional event. Let us imagine that this correlation is very weak and "causes" just a single binary number to occur instead of the second possible number (e.g. 0 instead of 1) which would have occurred in case of no external event. Then such a single flip would certainly not show up as a significant deviation of the RNG output - the effect would be not detectable. In our scheme however, such a weak correlation effect in principle can redirect the sensible non linear (chaotic) frequency walk of the first oscillator in such a way ("butterfly effect") that a resonance condition as described above is hit and a measurable deviation is generated in due course.

Fig. 2 shows the actual experimental arrangement in our Köln laboratory (Hagel & Tschapke, 2004). Actually our experiment consists of a sequential configuration of three samplers where the first sampler is triggered by the first oscillator (left of Fig. 2). The second sampler is triggered by a "1" output of the first sampler and the third sampler is triggered by a "1" output of the second sampler. The "1" outputs of the first and second sampler are coded as short pulses of light so that the three samplers are coupled optically like the first oscillator and the first sampler. The frequency of 120 Hz for the second oscillator linked to the first sampler has been chosen to obtain an average frequency of resonant departures from randomness of 48 hours obtained from simulation (Hagel, 2004). From this description it follows that the output of the first sampler has a balanced probability distribution for "0" and "1" (P(0) = P(1) = ½) while the second and third output sequences are unbalanced meaning that P(0) = ¾ for the second output and P(0) = 7/8 for the third output. The frequencies of the second oscillators connected to samplers 2 and 3 are running at 60 Hz and 30 Hz because the average output frequency of "1" at the first sampler is equal to f_1 / 2 and the average output frequency of "1" of the second sampler is f_1 / 4. In this way we arrive at the same average frequency relations between the input signal and the frequency of the second oscillators for every of the three units. The three sequences of random numbers and the exact time of their occurrence are constantly registered on a PC. From these explanations it becomes clear that the second and third sampler are dependent from each other in the following way:

Output of the first / second sampler = 0 implies Output of the second / third sampler = 0

Output of the first / second sampler = 1 implies Output of the second / third sampler = 0 or 1

The reason for using three instead of only one unit is to test the following idea:

If there exists a correlation between RNGs and external events, then unbalanced binary RNGs tend to correlate better with events of long time scales than with events of short scales.

This is supposed to be due to the fact that the variation with time of the output sequence of a binary balanced RNG is larger than the one of a unbalanced one.

The intention of the experiment is to investigate if there exist correlations between the behavior of the binary number sequence generated by the described apparatus and emotionally charged events in the environment of this unit or if these correlations do occur equally strong for distant events.

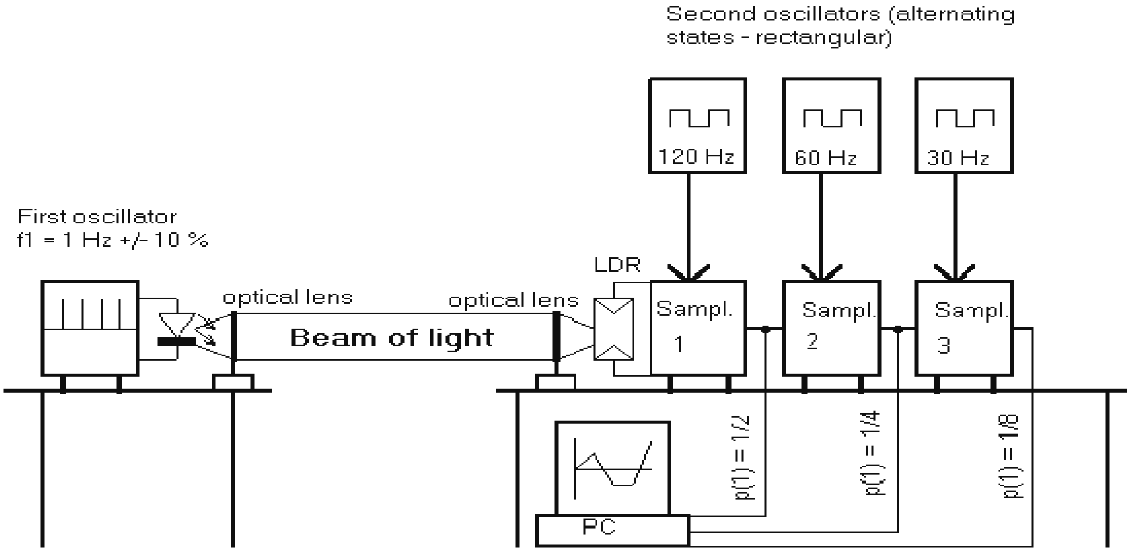


Fig.2: Schematic view of the LED experiment in the IPP Köln

In the experiment described we intend to answer to the following two questions:

1. With the single unit as described above is it possible to observe correlations between emotionally charged events and transitions between random behavior and nonrandom excursions of the system?
2. Can we find a certain local effect? This means to answer the question if the physical distance of the events from our experimental unit do influence the measured results.

Finally it should be clearly noted that *we understand our contribution as an exploratory study undertaken by our Köln laboratory*. The preliminary results described in this paper can therefore possibly be used for the generation of hypothesis rather than to test already generated ones. Testing such a hypothesis still has to be undertaken by subsequent investigations. Since to our opinion the scientific investigations of 'anomalous' effects is too much biased on non local concepts, we explicitly favored the investigation *of local influences*, where *local* means spatial vicinity of source and receiver.

METHODS

Generation and registration of the binary sequences

The unit described in the introduction runs with short interruptions since March 2003 and the binary sequences obtained are stored on one separate file for each day. As an example we show the first few lines of such a file:

Date: 5. 6.2003
Starting time: 21:20: 0

1	1	1	0	0.435483	0.435483
2	0	0	0	1.068499	1.503983
3	0	0	0	1.031184	2.535167
4	1	1	0	1.058888	3.594055
5	0	0	0	1.043781	4.637836
6	0	0	0	1.120108	5.757943

The first column contains the number of light pulses from the low frequency oscillator. The second, third and fourth columns represent the three binary outputs from the balanced and the two unbalanced sequences. The fifth column indicates the time interval (in sec) between the last and the present impulse (we use the mounting flank of each impulse as a reference). Finally the last column indicates the total time elapsed at the mounting flank of the present impulse.

At the end of each day (=24 hours after opening a one day file) the files are closed and copied automatically to a diskette. Then the results are evaluated using software on a separate computer in our laboratory.

Evaluation of the results

The evaluation procedure is done in the following steps:

1. By using appropriate Software Tools we compute the cumulative differences along the day first by fixing the beginning of the file as reference (CD(0) = 0). This is done for all three sequences of binary numbers. In addition the period of the low frequency oscillator is tracked as function of time as well. Second, if a possible deviation of the CD's starts at a given instant of time we also can fix the beginning of the Z=2 parabola to this instant of time as it is frequently done in the GCP data evaluation.
2. The cumulative differences for the three sequences of numbers are tested and inspected for deviations from randomness. Following the usual method we look for deviations of more than 2 standard deviations (σ) from expectation assuming that the random part of the binary sequences follows a Bernoullian distribution.
3. We check for correlations between the occurrence of departures form randomness and global as well as local events of which we assume that they are of emotional importance for many people. To be precise, we check as well for events that could be previewed (like big sports or religious events) as for non predictable events like catastrophes, crimes, accidents etc.
4. Finally we check for an experimenter effect. By this we mean a possible correlation of the unit with personal emotional states. In this case we put weight to strong emotions that are localizable in time to high precision like emotional reactions to good or bad news.

RESULTS

General remarks

It is evident that testing a hypothetical indicator for emotionally charged events inside and around a big city like Köln in the center of an highly populated industrial zone is not a trivial task specially if the mechanism for the assumed correlation is still unknown. We have to take into account that emotionally charged events take place at all times. Since we do not know how emotional states can interact with chaotic or random systems we may only apply heuristic models about such a type of anomalous interactions. Hence we assume that emotional events synchronously concerning a large group of individuals cause a stronger effect and will more probably correlate with a transition from random to resonant behavior of LED. On the other hand we can regulate the sensitivity of LED (meaning the average frequency of transitions between chaotic and resonant strings) by adjusting the frequency relation between the two oscillators. The results shown in the following chapter have been obtained with a frequency relation of 1:120. With this relation we obtained statistical curves in about 50 % of all days. This number agrees well with the result of computer simulation which we applied to the two oscillator system of LED (Hagel, 2004). For the remaining half we observed significant departures from randomness which to 50 % could be supposedly correlated to emotional events concerning the local population. For the remaining 50 % we could not attribute an event but it is practically certain that – considering the low manpower of our institute – we simply overlooked a considerable fraction. On the other hand we only observed a low fraction of important events concerning the local population which did not take place synchronous to a measurable deviation from randomness.

Sports events and their correlation with LED

There are few events that can trigger synchronous states of emotion in such an effective way like sports and specially football. Hence it is obvious for an experiment like LED to choose football matches as a possible target. They have the advantage to be announced in advance and it is absolutely certain that a big number of people (40000 – 60000) are present in the local arena. The following figure shows a typical plot of cumulative differences as produced by LED during an important local game:

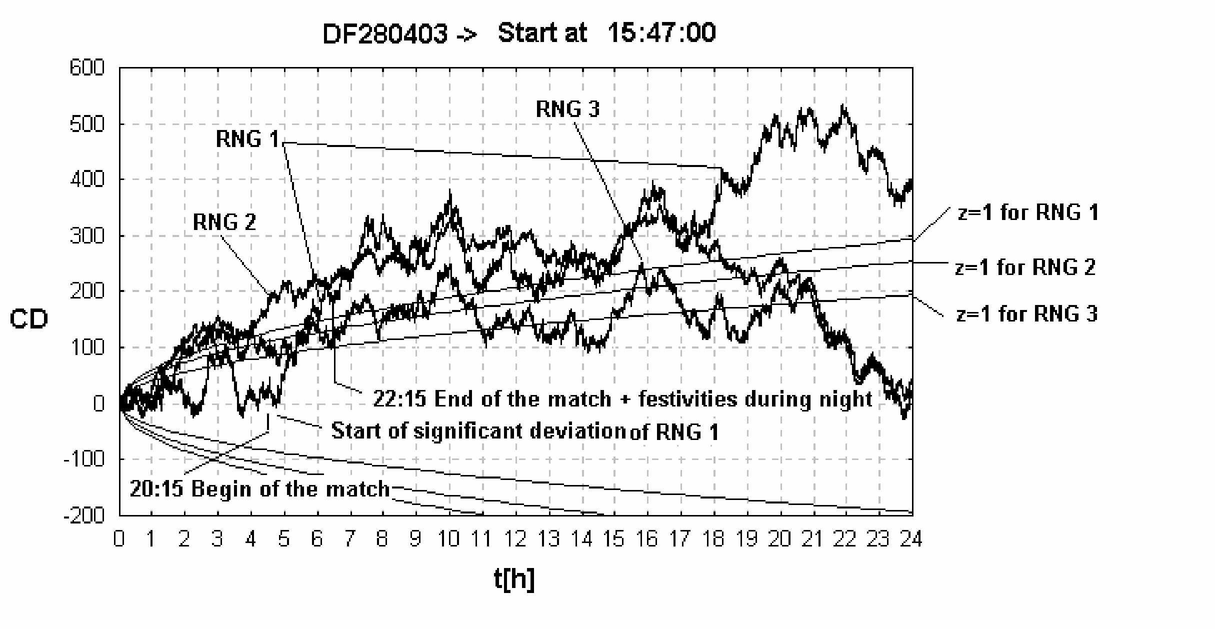


Fig. 3: Cumulative Differences for the three output sequences of LED during an important football match in the Rhein Energie arena of Köln

The abscissa shows the number of hours passed (0-24) since the starting time (15:47) of the current day (28-th April 2003). A transition from random to resonant behavior is visible at 20:37, a few minutes after the start of a very important match of the local football team, the FC-Köln. It was this match which decided about upgrading the team from the second to the first league of the German mastership and it was won by the FC! The observed deviation of the first sequence (balanced probabilities for "1" and "0" – green curve) reached a deviation of about 1.5 σ at the end of the match and continued to rise to 2σ for more than 1 hour. In fact, thousands of people after the match formed a large procession of a length up to 3 km singing and walking through the city of Köln. The CDs of the unbalanced sequences show a similar but much less pronounced behavior. The second strong deviation starting at 8:15 on the following day could be interpreted as a collective emotional reaction to the morning news about the FC-Köln having achieved the upgrade.

It has to be mentioned that *only local matches or matches of local importance* that have been transmitted in television showed *considerable effects*. Matches of different teams outside Köln did not significantly influence the CDs of LED.

Local Reaction to a spectacular plane accident in Africa

As a further example (in favour of the existence of locality) we show the CDs of LED on the 8 May 2003. Between 19:00 and 22:00 (local time of Köln) there happened a tragic accident near Kinshasa. The door of a plane opened a few minutes after take-off in 2000 meters above ground. As consequence 250 passengers have been thrown out of the plane by the stream of air and died.

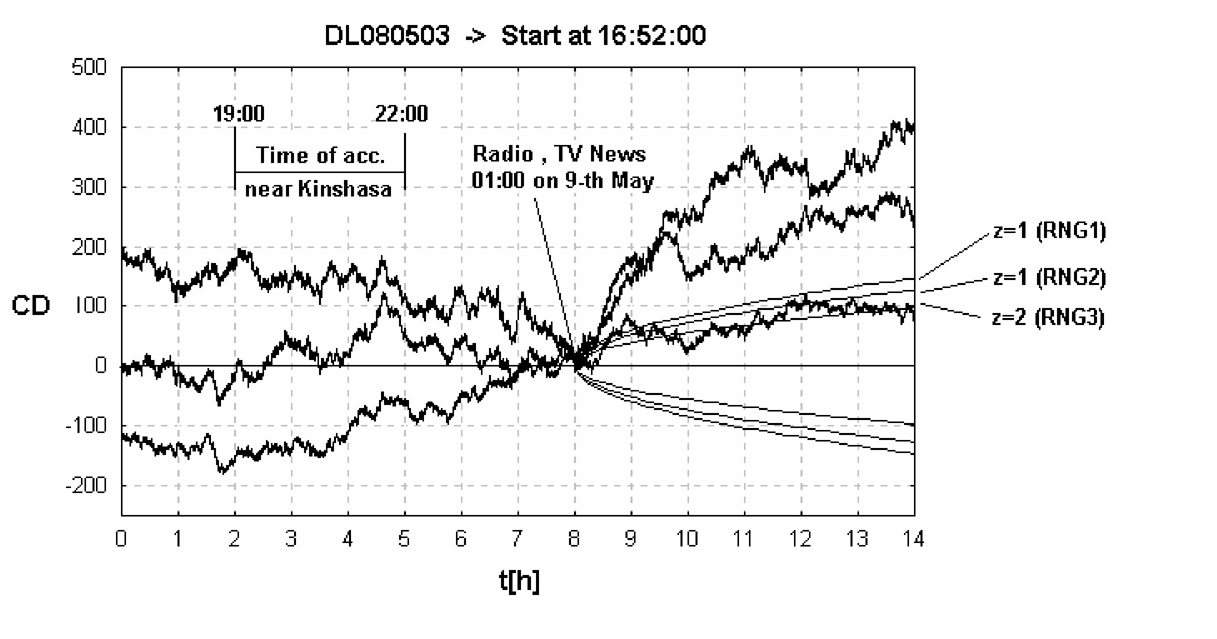


Fig. 4: The CDs of LED during and after a serious plane accident in Africa

As can be seen, the event itself which must have generated very strong local emotional states near and in Kinshasa did not generate significant reactions of LED. However at the instant of time of first news about the accident reaching Germany (and thus Köln) we see a strong reaction. As before, the balanced oscillator shows the most distinct reaction.

The experimenter effect

The Institute of Psycho Physics (IPP) is located in the house in which one of the authors (M. Tschapke) lives her private life and runs the experiments. If the assumption of a correlation between emotional states and the behavior of LED is correct as well as the one of a dependence on the physical distance of emotional states and our experiment, then we must be aware that also personal emotional events concerning the experimenter can correlate with LED. In this sense the experimenter and the experiment form a closed unit. In fact we found several (5-10) cases in which such a correlation appears with very high precision. Up to today such results occurred exclusively in the case of an extraordinary strong emotional movement of the experimenter. Fortunately for some of these events we have very precise information about the instant of time of occurrence (phone calls or personal messages) so that a detailed analysis was possible. In Fig.5 we show one such example. On March 17-th, 2003 exactly at 13:20 some news of very emotional content reached M. Tschapke from our institute: At this instant of time her brother informed her that his wife will have a baby. M. Tschapke described her reaction by “a very strong positive emotion setting in rapidly and lasting for a few hours”.

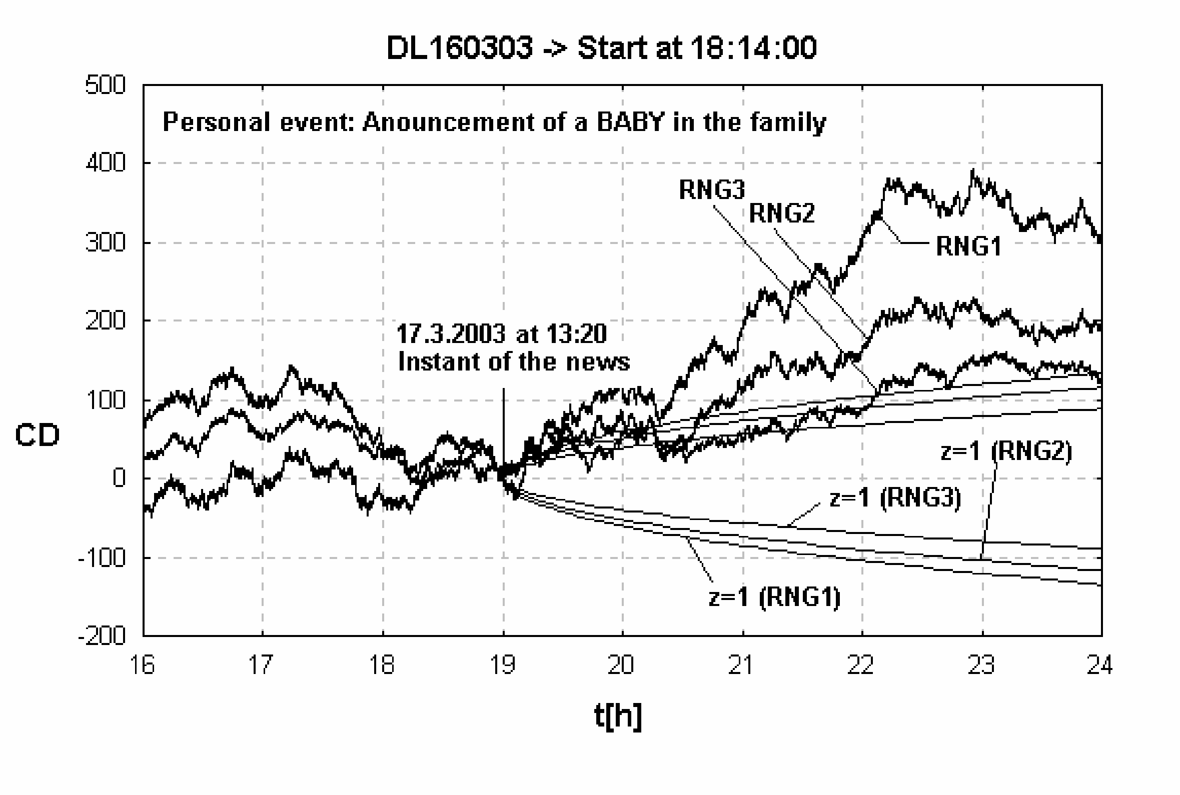


Fig. 5: Strong deviation of LED at the instant of time of a emotionally charged news

Based on this information we decided to investigate this case in more detail: In order to understand the rapid excursion at 13:20 we plotted the frequency of the first oscillator as function of time over an interval of several hours before and after the event (Fig. 6).

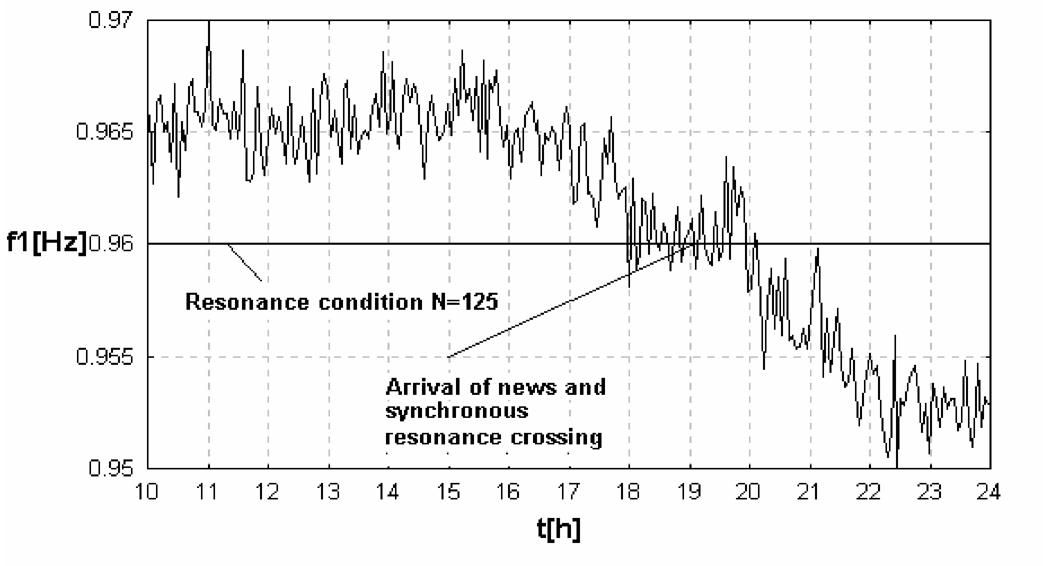


Fig. 6: Frequency of the first oscillator as function of time

Every point in this figure corresponds to the averaged frequency over the 200 last periods of the first oscillator. While the average frequency was nearly constant for at least 6 hours (10–16 hours after start), a sudden descend of the frequency sets in three hours *before the instant* of the good news. As a matter of fact the center of crossing of the frequency curve with the resonance condition

$$125\ f_1 = f_2 = 120\ \text{Hz} \Rightarrow f_1 = 0.96\ \text{Hz}$$

appears *exactly* at the critical instant and in this way causes the observed deviation. This can be understood at least as one possible mechanism of correlation between emotional events and random processes.

CONCLUSIONS AND DISCUSSION

A new type of binary number generator (LED) based on a two oscillator system has been described. It consists of a low frequency electronic oscillator with chaotic frequency variations and a higher frequency oscillator generating symmetric rectangular oscillations. The output sequence consists of interleaved strings of random number sequences and resonant sequences leading to departures from randomness. We investigated the question if transitions between random and resonant behavior are related to local emotionally charged events in the city of Köln.

Since the beginning of the long term experiment we found that strongly emotional events in and around Köln (like sports events) do occur synchronous to observed transitions of LED.

It is clear that the hypothesis about a local character in the correlations of the two oscillator system and emotional events is still speculative. At the time we have too few supporting data from LED to state more than a certain doubt about an exclusively non local nature of such correlations.

However, there are some observations that might indicate that there exists a certain local component:

1. Important sports events outside of Köln - e.g. in Essen (80 km) or Dortmund (120 km) - did not show up in our results even if they were of emotional importance to the population of theses cities.
2. Events like severe accidents outside Köln and even Europe did show up at the occasion of the news about these events reaching Germany (and Köln). The most impressive example of such an event is shown in Fig. 4.
3. Severe accidents in the local area of Geneva (Switzerland) – the second seat of our institute – which did not make their way into international news distribution did never show any effect on LED.

It is also evident that there exists an alternative hypothesis which could explain these "local effects": Large events like football matches or similar imply various physical effects in the neighborhood of their occurrence. These are e.g. propagation of micro-seismic waves or electrical perturbations due to a large local energy consumption etc. These are clearly local effects, they can in principle affect the LED arrangement and they all follow an inverse distance law. So they could in principle account for an explanation of 1. It will however not be so evident to explain the delay effect described in 2. since news transmissions about distant serious accidents will not generate seismic waves like a local match and no significant changes in local energy needs. Evidently we tried everything in the range of our possibilities to avoid external physical influences to act on the two oscillator system: The first oscillator is enclosed into a hermetic metallic box to avoid the impact of external electrical fields. All oscillators and electronic equipment are mounted on massive tables by using gum feet with small cross sections in order to damp seismic vibrations. Finally *every* single oscillator has a *separate electrical power supply* of which all are stabilized voltage sources of improved quality.

It was a striking observation that even strong emotional events occurring to the experimenter seem to show up in the cumulative differences of the binary output sequences. However, the critical question must be asked if these deviations are caused by a local influence of the experimenter or by a more general non local influence just through his expectation on the results of the entire experiment. This seems to be a very fundamental question in the whole GCP (and LED) studies and should be discussed in more detail: Evidently if we investigate a new and unknown type of interaction it becomes difficult to distinguish between an effect independent of the experimenter and one caused by him. This is equally true for the GCP/LED studies as well as for all kinds of experiments aiming for the understanding of paranormal information transfer (telepathy etc...). We believe that one important factor in favor of a local effect is the *unexpectedness* of the events under consideration. It is certainly correct that messages like the announced birth of a baby are received in any part of a large city every day. And many of these messages reach their targets in an unexpected way. Hence if the assumed correlation effects in LED belonging to such messages were primarily non local would this not mean that we should observe deviations nearly at any instant? On the other side there is also an important argument in favor of a non local behavior: As pointed out in the abstract the average number of deviations in time agrees well with the expected value from computer simulation of the two oscillator system. It seems that only the temporal arrangement of the deviations is "chosen" by the system in such a way that the observed agreement between external emotional states comes together.

Independent of the final solution to this seemingly contradictory problem we had the chance to study these coincidences in much detail and we could identify resonance crossings caused by a slow frequency walk of the first oscillator to play an important role. In this way we could uncover the basic mechanism of the observed transitions between random and resonant behavior of LED.

However, it remains still unclear by what property of the system such resonant crossings in random systems *appear synchronous to emotional events*. An answer to this question would probably help to explain the phenomenon of synchronistic events described by C.G. Jung and in how far this fascinating and exciting effect is of non local or local nature.

References

Hagel, J. (2004). Numerische Simulation eines chaotischen Zwei Oszillatoren Systemes ohne physikalische Koppelung, Publication in preparation.

Hagel, J. & Tschapke, M. (2004) Der Local Event Dedector (LED) Ein neue experimentelle Anordnung zum Hinweis auf Korrelationseffekte unter dem Auftreten lokaler emotional geladener Zustände, submitted for publication to *Zeitschrift für Anomalistik*

Lichtenberg, A. J. & Liebermann, M. A. (1983). *Regular and Stochastic motion,* Springer Verlag, New York, Heidelberg, Berlin.

Nelson, R. D. (1997). *Multiple field REG/RNG recordings during a global event.* The electronic Journal for Anomalous Phenomena (eJAP), 1997. http://www.psy.uva.nl/eJAP

Nelson, R. D. (2003). *Private Communication*

Address for correspondence: Johannes Hagel,
Institut für Psycho-Physik (IPP), Stüttgerhofweg 6B,
50858 Köln Junkersdorf, Germany.
E-mail: cephir@netcologne.de

ANOMALOUS FLUCTUATION OF RNG DATA IN NEBUTA: SUMMER FESTIVAL IN NORTHEAST JAPAN

Tatsu Hirukawa & Masato Ishikawa
School of Information and Communication, Meiji University, Tokyo

ABSTRACT

This paper describes a pilot study investigating the mind-matter interaction during the Japanese traditional festival. On August 2nd and 3rd, 2003, fluctuation of RNG was measured in the first night of *Aomori-Nebuta*, one of the most popular folk festivals in northeast Japan. In 6-day festival, around 600 thousand people gathered every *nebuta* night and enjoyed singing and dancing with rhythmical drums, like carnivals in Latin American countries. Authors expected to find some significant deviation of the field RNG in this festival, particularly because some *yakuza* (Japanese gang)-like groups called *karasu* (crow) become so excited that they sometimes turn into violent mobs, that is, mass emotional excitement may effect the outputs of RNG. The hypothesis that the deviation in the RNG data appears during the festival period was tested by the cumdev and windowing methods. By the cumdev method, significant deviation was obtained in the overall accumulation of chi-square values during the 110-minute festival period (18:50-20:40, Aug. 2nd): Z^2-1=200.1, p=0.041 (one-tailed; *df*=6600). However, accumulation of Stouffer Z value during the same period showed no significant deviation: Z=1.00, p=0.32 (two-tailed). In the windowing method, three different sizes of window width (15, 30, and 60 minutes) are used for analyzing the 19-hour data from 11 am (Aug. 2nd) to 7 am (Aug. 3rd). Each highest peak of window-shifting accumulation of chi-square value in the three different windows was found in the festival period, which is significant: in 19:45-19:59, Z^2-1=106.5, p=0.0075 (one-tailed; *df*=900); in 19:30-19:59, Z^2-1=140.0, p=0.0111 (*df*=1800); in 19:00-19:59, Z^2-1=202.6, p=0.0093 (*df*=3600). All the ends of peak windows are the same (19:59). On the other hand, each highest peak of the absolute value of Stuffer-Z in three different windows was also significant but found at the end border of festival period: in 20:45-20:59, Z=-2.409, p=0.0160 (two-tailed); in 20:30-20:59, Z=-2.187, p=0.0288; in 20:00-20:59, Z=-2.355, p=0.019. Just as the result of chi-square deviation, all of the ends of highest windows are the same (20:59). The overall results suggest the field RNG analyses on local festivals are promising. More data are required for proving the hypothesis and discussing correlates between the RNG deviation and a variable such as the number of participants.

INTRODUCTION

This paper describes a pilot study investigating the mind-matter interaction during the Japanese traditional festival. Its aim is to check if the field RNG analyses on local festivals are promising or not.

Field RNG

The previous researches suggest that the mind-matter interaction may occur, when a large number of people form a shared emotion, a shared attention, or a specific state of consciousness. The researches are referred to as FieldREG (Nelson et al., 1996; Nelson et al., 1998) or field consciousness (Radin, 1997; Radin, 2002). The mind-matter interaction seems to be found as deviation on a random number generator (RNG), or a random event generator (REG), in which a logical XOR (eXclusive-OR) guarantees zero deviation in the long run (Ibison, 1998; Nelson, 2001).

Significant RNG deviation was often reported during some meetings (Nelson et al., 1996), broadcasting events (Radin et al., 1996), and sport events (Bierman, 1996). The Global Consciousness Project (Nelson, 2001) built a worldwide network of RNGs connected by Internet. The authors keep an RNG site running, which is the only site of GCP in Japan as of today. The GCP data showed significant deviation during some global events such as New Year, World Cup Soccer, and the Olympics. The most prominent results were obtained by the data of September 11, 2001 (Radin, 2002; Nelson, 2002).

To reveal the properties of mind-matter interaction, further research results on local events as well as global events are required, because a distinct shared emotion and attention are found in different cultural backgrounds.

Particularly, indigenous festivals and religious rituals are important topics, because participants of such ceremonies usually share a special state of consciousness and may experience psi-related phenomena. Nelson et al. (1998) review some results measured in such situations as healing rituals of native American and "paganist" ceremonies. In Japan, many kinds of festivals and rituals have been held in the long tradition of the complex syncretism of native animism, shamanism, Shintoism, and Buddhism. Even today, this syncretic worship for gods, spirits, and ghosts are popular among Japanese people. Yoichi et al. (2002; 2004) gathered RNG data focusing on Japanese New Year festivals, and Kokubo et al. (2002; 2004) researched "haunted houses" (i.e. RSPK cases) in Japan. They reported some significant results.

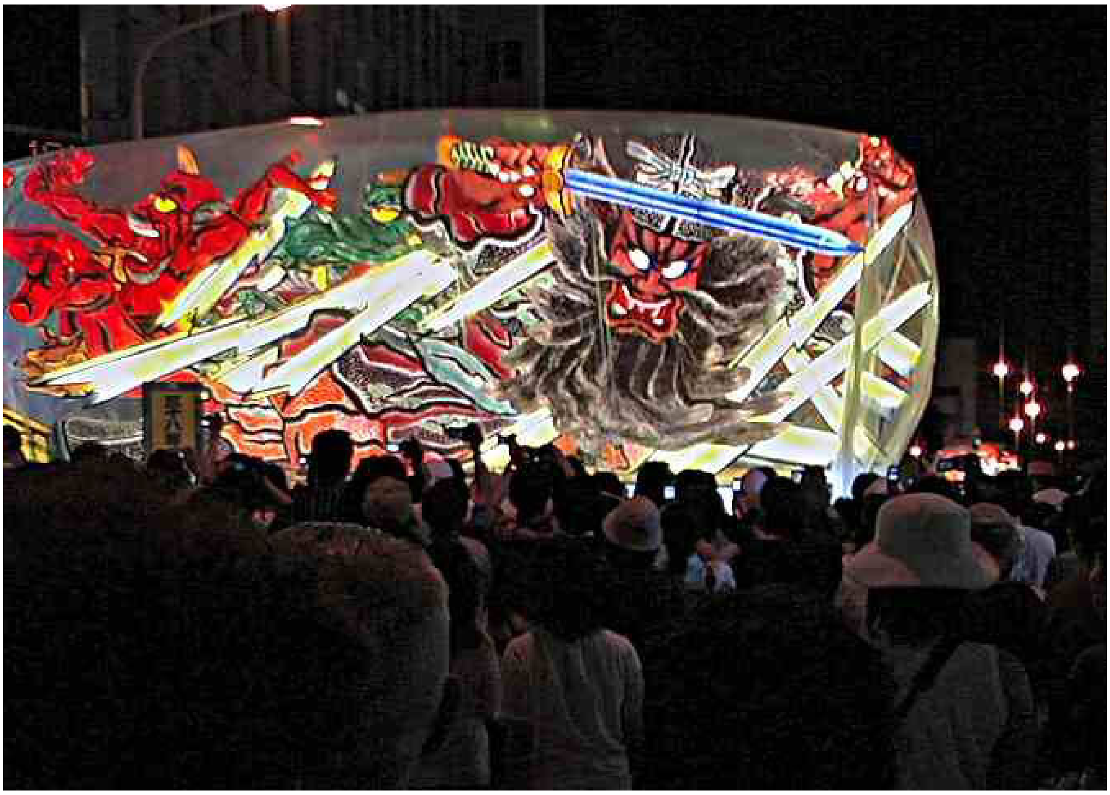


Fig. 1: A *nebuta* of the thunder-god swinging his sword. Each *nebuta* features fighting *samurai*s or rough gods representing awful power of the nature. All *nebuta*s were covered with plastic sheets because it was drizzling at the first night of the last festival. (Aomori, Japan)

Nebuta Festival

Aomori-Nebuta held in Aomori (140:44' E, 40:49' N), a northernmost city in mainland Japan, is one of the most popular summer festivals in Japan. During 6-night festival in every summer, *nebuta*s, huge illuminated figures representing local gods and legendary *samurai*s (Fig. 1) go round the streets in central Aomori city. Like the carnivals in Latin American countries, 2000 *haneto*s (jumping [wo]men) dance around each of a dozen of *nebuta*s with repeating 4-beat rhythm: "lasseh-lah", and people release their "energy repressed for a year" – particularly for long gloomy winter. Many people also come to join the festival from other cities and even from foreign countries to be not only spectators but also *haneto*s. Last summer, around 600 thousand people gathered at the center of Aomori city every *nebuta* night (around from 7 to 9 pm). During and after the festival, people stop in stalls and bars to buy souvenirs, to eat and to drink. Most people go back to their home until around 11 o'clock in the midnight before the last train leave Aomori station and bars and restaurants close. Few people wander around the street for overnight. However, during the festival, some *yakuza* (Japanese gangster) -like group of *haneto*s become so excited that they sometimes turn into violent mobs. They are called "*karasu* (crow) *haneto*", or simply "*karasu*" because of typically wearing black *kimono*, and regarded as a social problem in the area (Anami, 1999) (Fig 2). Although the origin of the festival is not clear, people have celebrated their short summer for hundreds of years. According to the historical record, many people have been injured and even killed during the festival of every summer.

Fig. 2: "NO! Just say NO!! Stop the *KARASU*! If you play *karasu*, you must be punished by the law! Last year, 5 people were arrested!" Every summer, we can see many warning posters distributed by Aomori local police. *Karasus* typically ware black or *yakuza* (Japanese gang) -styled *kimonos* or *tokko* (*kamikaze* fighter) clothes. (Aomori, Japan)

METHODS

Equipment

A portable computer controls Orion (http://www.randomnumbergenerator.nl/), where a sequence of random numbers are generated and transmitted through its RS232C port. Sampling software FRED (see *acknowledgments*) records 200-bit random numbers per second; every bit (0 or 1) is summed up for a second to make a value N whose MCE is 100 and VAR is 50. Each number N is normalized to Z score: Z = (N-100) / sqrt(50). The Z scores are analyzed using MS-Excel and statistics software SPSS.

Measurement

Hirukawa, the first author, measured RNG data during the last *Aomori-Nebuta* festival on August 2nd 2003, the first day of the 6-day festival. On the day, *nebuta*s went round the center of Aomori city around from 18:50 to 20:40 (Japan Standard Time). RNG data were recorded from 11 am of 2nd to 7 am of 3rd in a room of a hotel located about 300 meters apart from the city center where the *nebuta*s went around. There was no person in the room during the festival to minimize the experimenter effect, although Hirukawa slept in the room from 0 to 7 am of 3rd. No data were recorded from 23:00 to 23:59 due to check the computer.

No special event or accident occurred in the period other than the festival although the stuffs prepared for it from the morning. No one was killed or severely injured during the festival of the night under the strict control by local police (Anami, *pers. comm.*).

Hypotheses

The only explicit hypothesis we posit is the same as other researchers' (e.g., Bierman, 1996; Yoichi et al., 2004).

> The deviation in the RNG will appear during the festival period while the RNG will operate randomly during the other periods, where the deviation means that significantly more **1**s/**0**s are generated than the mean chance expectation. The deviation is specified by the combination of Z scores, each of which is calculated by the occurrence of **1**s and **0**s in a second.

Analyses

To test the hypothesis shown above, we employed two types of analyses: *cumdev* and *windowing* (Radin et al., 2003). The cumdev method evaluates the deviation, accumulating per-second Z scores over the whole period. The windowing method compares the deviation in smaller width of time windows. In both methods, each per-second Z score is combined into longer periods of time: (1) as a change in variance in the form of an N-time sum of Z^2-1 (MCE=0; VAR=2N), which corresponds to the chi-squared distribution with degrees of freedom (*df*) equal to the number of seconds, N, or (2) as a mean-shift in the form of Stouffer Z, which corresponds to the simple combination of Z scores.

RESULTS

Cumdev

The fluctuation of RNG outputs during the festival is analyzed by the cumdev method. Its purpose is to check if any significant deviation is found during the festival period in comparison with the mean chance. The following two line graphs show cumulative deviation of chi-square (Fig. 3) and mean-shift (Fig. 4). Each graph accumulates three sets of 110-minute data: during the festival period (18:50-20:40 JST), just before the period (17:00-18:50 JST), and just after the period (20:40-22:30 JST). The before-festival and after-festival data are treated as controls.

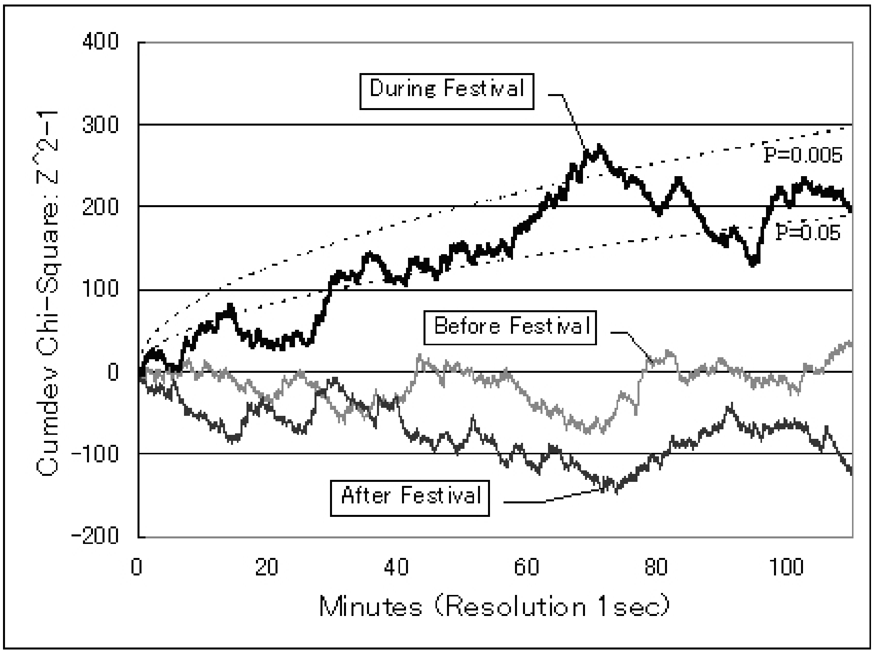


Fig. 3: Cumulative chi-square deviation of RNG data during and around *Nebuta* festival

In Fig. 3, the cumulative chi-square curves are shown. The y-axis is N-time sum of Z^2-1, where N corresponds to the number of seconds and *df*. The dotted lines are the border of significant p values, which is calculated by the approximation of standard distribution whose MCE is 0 and VAR is 2N. The during-festival curve goes up to the highly significant level at the middle of the festival period; the overall value in the 110-min period is Z^2-1=200.1, p=0.041 (one-tailed; chi-square, *df*=6600). The control curves are not significant.

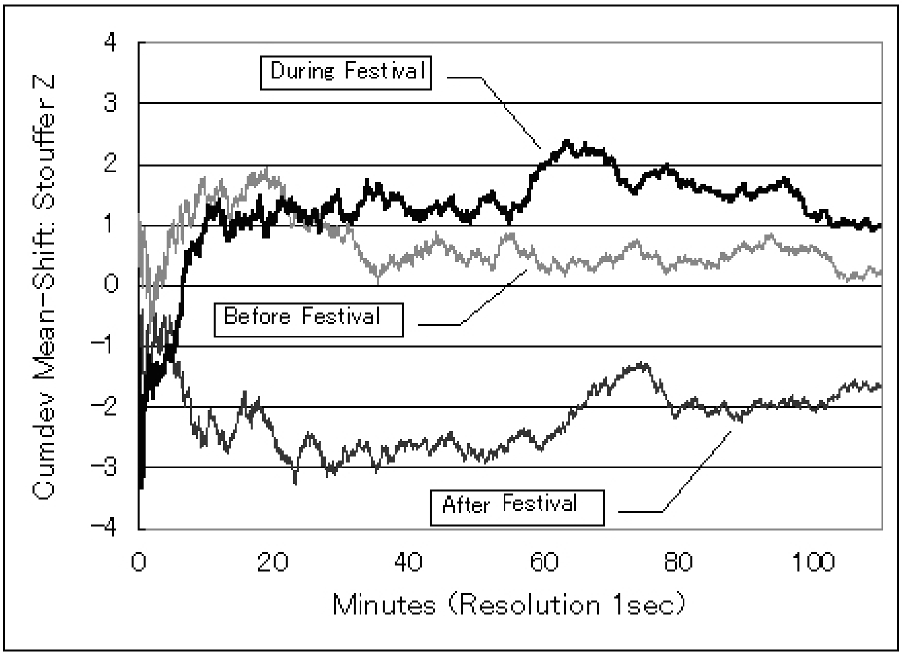


Fig. 4: Cumulative mean-shift of RNG data during and around *Nebuta* festival

In Fig. 4, the cumulative mean-shift curves are shown. The y-axis is Stouffer Z, the simple combination of Z score of each second. The during-festival curve shows that more **1**s were generated in the former half of the festival period than **0**s, while less **1**s were generated in the latter half. The overall value in the 110-min period is not significant: Stouffer Z=1.00, p=0.32 (two-tailed). The control curves are not significant as well.

Windowing

The deviation of RNG outputs is analyzed by the windowing method. Its purpose is to check if any significant deviation is found during the part of festival period in comparison with the other periods. The following two bar graphs show changes of chi-square (Fig. 5) and mean-shift (Fig. 6). Three different sizes of windows are arbitrarily decided and used in this analysis: 15, 30, and 60 minutes. For example, the left-most bars are the accumulation from 11:00:00 to 11:59:59 in 60-minute window, from 11:00:00 to 11:29:59 in 30-minute window, and from 11:00:00 to 11:14:59 in 15-minute window. There is no data from 23:00 to 23:59 because of the computer check.

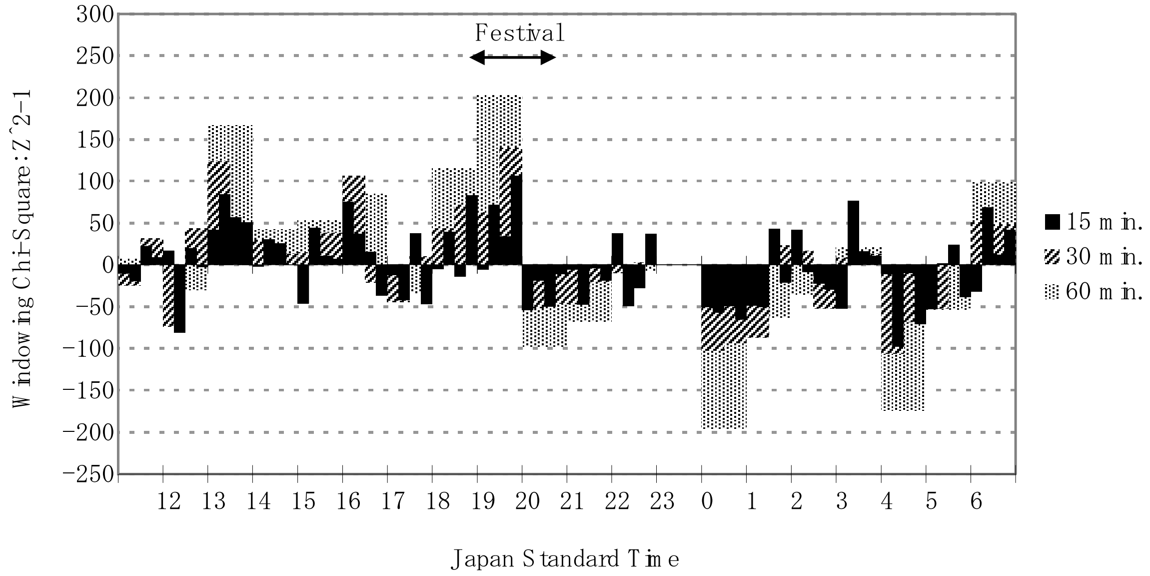


Fig. 5Chi-square deviation of RNG data in 15, 30, and 60-min windows

In Fig. 5, although some bars show significant deviation not only within the festival but also in other periods, each highest bar of three different windows is found within the period of the festival (18:50-20:40). All peaks are significant in 5-percent level, and all the ends of peak periods are the same (19:59) [see Table 1]. After 20 o'clock, the latter half of the festival, chi-square values go down below zero.

Table 1 Accumulation of Z^2-1 values of the highest bars in 15, 30, and 60-min windows

Window	Time	Sum (Z^2-1)	*df*	p (one-tailed)
15 min.	19:45 - 19:59	106.5	900	0.0075
30 min.	19:30 - 19:59	140.0	1800	0.0111
60 min.	19:00 - 19:59	202.6	3600	0.0093

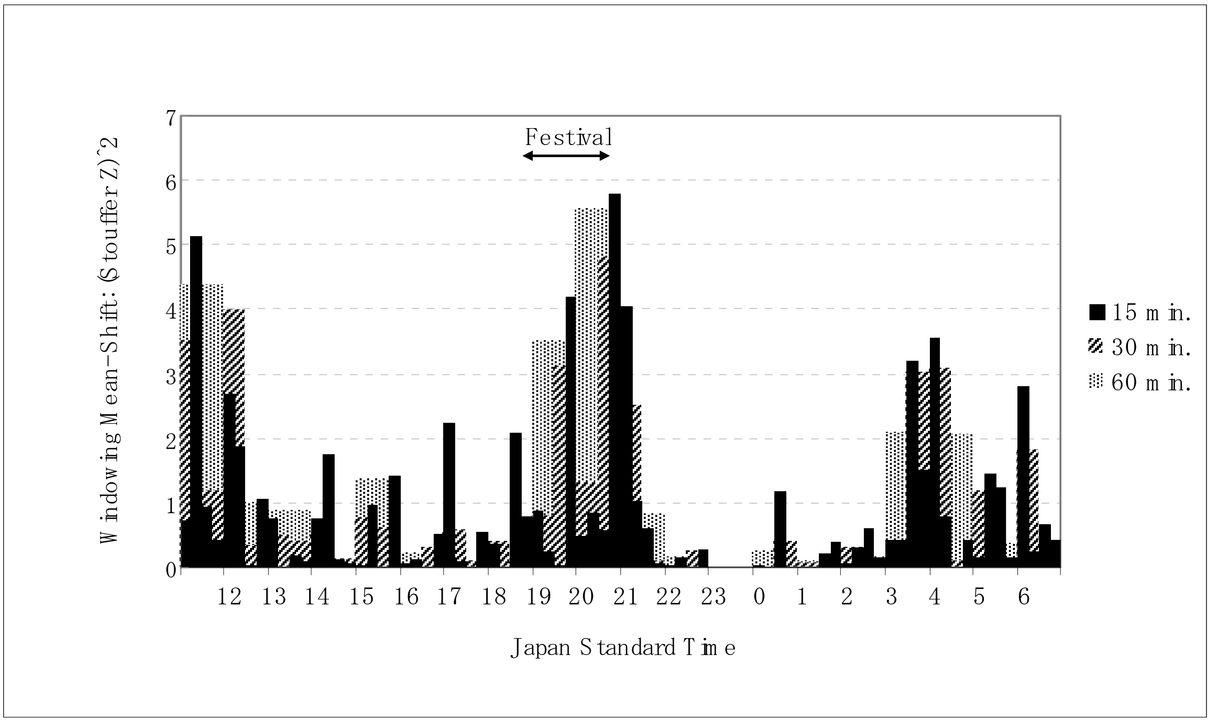


Fig. 6: Mean-shift of RNG data in 15, 30, and 60-min windows

In Fig 6, just as the result of chi-square deviation, each highest bar of three different windows is significant in 5-percent level and all of the ends of peak periods are the same (20:59) [see Table 2]. However, this result is marginal, because the festival itself officially ended at 20:40. The highest value of 15-min window is found just after the festival.

Table 2: Stouffer Z and (Stouffer Z)^2 values of the highest bars in 15, 30, and 60-min windows

Window	Time	Stouffer Z	(Stouffer Z)^2	p (two-tailed)
15 min.	20:45 - 20:59	-2.409	5.803	0.0160
30 min.	20:30 - 20:59	-2.187	4.782	0.0288
60 min.	20:00 - 20:59	-2.355	5.545	0.0185

CONCLUSION AND DISCUSSION

Significance of the results

The chi-square analysis resulted in the significant deviation during the festival period in comparison with the mean chance and the other periods. The cumdev method showed significant chi-square deviation in the festival period (Z^2-1=200.1, p=0.041). The windowing method located the most deviated chi-square peaks at the middle of festival period, when using 15, 30, and 60-min windows.

The mean-shift analysis resulted in no significant deviation during the festival period, but found some characteristic deviation just after the festival period. The cumdev method did not show any significant

Stouffer-Z deviation in the festival period (Stouffer Z=1.00, p=0.32). The windowing method located the most deviated Stouffer-Z peaks at the end of festival period, when using 15, 30, and 60-min windows.

Overall the results are partly affirmative to the hypothesis, but to prove it, more field RNG analyses on some local festivals should be repeated. The results suggest such analyses are promising.

Future Research

We found much deviation of Stouffer Z in the latter half and just after the festival, while the accumulation of Z^2-1 changed to negative values in the same period. This seems a paradoxical result. However, we don't have enough data to generalize it as a common trend in this kind of field RNG experiments. More data are needed for further discussion such as correlation between the RNG deviation and the number of participants, or the types of emotion etc. *Nebuta* is not a very sacred ritual related to any religious cult, but a festival in folk society. We have already collected and are now analyzing the RNG data of other religious festivals in Japan, such as *Osorezan-Taisai*: a yearly ceremony to appease dead spirits by Buddhist monks (northeast Japan) and *Onbashira-Matsuri*: a ritual for the renewal of guarding wood pillars around *Suwa* Grand Shrines (central Japan).

ACKNOWLEDGMENTS

Authors thank Dr. Dean Radin for allowing us to use a beta version of Field REG Device (FRED) code and Prof. Toru Anami, specialist of Nebuta folklore, for his helpful advice. We also thank Ms. Hiroko Hagiwara for her help.

REFERENCES

Anami, T. (1999). Aomori nebuta to karasu haneto (*Aomori nebuta* and *karasu haneto*). *Shukusai no Hyaku-nen (One Hundred Years of the Festivals).* Tokyo: Domes. (In Japanese)

Bierman, D. J. (1996). Exploring correlations between local emotional and global emotional events and the behavior of a random number generator. *Journal of Scientific Exploration*, 10, 363-373.

Ibison, M. (1998). Evidence that anomalous statistical influence depends on the details of the random process. *Journal of Scientific Exploration*, 12, 407-423.

Kokubo, H., Yamamoto, M. & Hirukawa, T. (2004). Case report on strange electro-magnetic signals: Research of "poltergeist" incident in Gifu, Japan. *Livro de Registiro de Trabalhos Apresentados, 2 Encontro Psi, Refletindo Sobre o Futuro da Parapsicologia (Proceedings of the Works of Presentators, 2nd Psi Meeting: Reflection to the Future of Parapsychology),* 66-75. Curitiba: Faculdades Integradas "Espírita"

Kokubo, H., Yoichi, H. & Yamamoto, M. (2002). Data analyses of a field number generator. *The Japanese Journal of Parapshychology*, 7, 11-16. (In Japanese)

Nelson, R. D. (2001). Correlation of global events with REG data: An Internet-based, nonlocal anomalies experiment. *Journal of Parapsychology*, 65, 247-271.

Nelson, R. D. (2002). Coherent consciousness and reduced randomness: Correlations on September 11, 2001. *Journal of Scientific Exploration.* 16, 549-570.

Nelson, R. D., Bradish, G. J., Dobyns, Y. H., Dunne, B. J. & Jahn, R. G. (1996). FieldREG anomalies in group situations. *Journal of Scientific Exploration*, 10, 111-141.

Nelson, R. D., Jahn, R, G., Dunne, B. J., Dobyns, Y. H. & Bradish, G. J. (1998). FieldREG II: Consciousness field effects: Replications and explorations. *Journal of Scientific Exploration*, 12, 425-454.

Radin, D. I. (1997). *The conscious universe.* San Francisco: HarperEdge.

Radin, D. I. (2002). Exploring relationships between random physical events and mass human attention: Asking for whom the bell tolls. *Journal of Scientific Exploration*, 16, 533-548.

Radin, D. I., Rebman, J. M. & Cross, M. P. (1996). Anomalous organization of random events by group consciousness: Two exploratory experiments. *Journal of Scientific Exploration*, 10, 143-168.

Radin, D. I., Taft, R. & Yount, G. (2003). Possible effects healing intention on cell cultures and truly random events. *Proceedings of the 46th Annual Convention of the Parapsychological Association*. 162-183.

Yoichi, H., Kokubo, H. & Yamamoto, M. (2002). Anomaly of random number generator outputs: Cumulative deviation at a meeting and New Year's holiday. *Journal of International Society of Life Information Science*, 20(1), 195-201.

Yoichi, H., Kokubo, H. & Yamamoto, M. (2004). Anomaly of random number generator outputs (II): Cumulative deviation at New Year's holiday. *Journal of International Society of Life Information Science*, 22(1), 142-146.

Address for correspondence: Tatsu Hirukawa,
School of Information and Communication,
Meiji University, 363 Kenkyuto, 1-9-1 Eifuku,
Suginami-ku, Tokyo 168-8555, Japan.
E-mail: hirukawa@tkk.att.ne.jp

THE POSSIBLE ROLE OF INTENTION, ATTENTION AND EXPECTATION IN REMOTE VIEWING

Joseph W. McMoneagle & Edwin C. May
Laboratories for Fundamental Research, Palo Alto

ABSTRACT

Joseph W. McMoneagle has participated in 44 on-camera demonstrations of remote viewing 35 of which would be considered as successful; that is, if they had been assed by the usual blind rank-order method, they would have easily been ranked correctly in first place. The question we address here is, "What, if anything, is special about these cases?" Under US Government funding, the research track record of what is known as STARGATE was exceptional. Perhaps the success could be attributed to the near exclusive use of highly talented special participants. However, we speculate here that the ill-defined concepts of *intention*, *attention*, and *expectation* were/are major contributors to the success of application-oriented, laboratory, and media-centered trials. As an illustration of these points, we provide a detailed description of the protocol and results of a recent demonstration trial conducted for the National Geographic Channel that was carried out in LFR's remote viewing laboratory in Palo Alto, California. The produces and staff of Pioneer Productions dedicated one individual for four days just to prepare for the shoot. Her duties were to learn about what constitutes a good remote viewing target, identify 6 targets within 25 km from the laboratory, prepare two sets of target packs, identify a neutral 3rd party individual to secure these materials, and act as beacon person during the trial. The full team included a camera crew of three and a single producer. At the time of the trial all people present with the viewer were blind not only to the individual randomly selected target but also to the complete six-fold target pack. The response was blind rank-order assessed on-camera, and the correct target was matched as 1st place. The qualitative correspondence with the intended site was excellent and typical of the 35 of 44 other media trials provided by McMoneagle. This single trial serves as an exemplar of an ideal application of *intention*, *attention*, and *expectation*.

INTRODUCTION

We provide a qualitative and first-hand account from part of a research team that has worked together for 25 years. We will focus upon remote viewing trials that were conducted under the glare of various television cameras. The first of these was conducted by LMNO Productions in California. They approached us to be on a segment for a US National show called Put to the Test which aired on the ABC network in November 1995.

Since that time, Joseph W. McMoneagle has participated in 43 additional, real-time, on-camera remote viewing demonstrations with national level television production companies in six countries. About 35 of these were of such qualitative agreement with the intended site, that if they had been analyzed by the usual blind ranking technique commonly used within the laboratory, they would have easily been matched as a first place.

In a series of Japanese targets filmed in five presentations, now spanning a period of two years, the majority of these targets have been missing people wherein only the names and birthdates were known. In some of these cases, the place and where the person was last seen was known, and intensive searches for these people by police and private detective agencies met with little success over five to seven years. Only the name and birth date information was brought to the filming of the remote viewings which took place in the State of Virginia. All information relative to the missing people was placed within sealed envelopes and Joe was required to provide a complete description of the person, their age, location, and how to find them in Japan.

Joe was targeted only with what was contained in the sealed envelope; that is, something like, "This envelope contains what is known about a missing person in Japan. Please provide information about that person and to the whereabouts of that person now." The remote viewing information was then taken back to Japan and provided to a detective agency to be used to hunt for the missing people, with the results later televised. Joe did not receive any feedback prior to the actual show depicting the results of these searches shown in Japan, in some cases many months after the remote viewing session.

Five out of seven missing people were located using the remote viewing material alone. All portions of the search and the use of the remote viewing materials in the search were clearly documented and demonstrated on film in the formal program as feedback. In the case of people who have been found, they had been missing from 7 to 30+ years. In two of the cases, it was acknowledge on Nippon TV that both the police as well as professional detectives were unable to locate the missing persons Joe had been able to locate from his dining room in Virginia while targeting a sealed envelope using remote viewing.

Two additional missing children cases have been turned over to the police with solid leads because they are considered now to be criminal cases as a result of the information developed by remote viewing and will no longer be used for viewing or media purposes in Japan.

Joe has been challenged three times by highly skeptical panel members participating on the program during the live broadcast, usually by their announcing they have had their picture taken somewhere in Tokyo on the way to the studio and demanding Joe draw the location, which he has done accurately on all three occasions. In the most recent example, he also nearly named it as well – "I want to call this a Sports Arena, but that's not what it actually is." It was actually – "Sports Arena Auto."

The latest US-based demonstration trial was for the National Geographic channel that was conducted in March 2004 to be aired sometime in the fall of 2004, and we will describe this example in detail

INTENTION, ATTENTION AND EXPECTATION

These terms, almost by definition, are imprecise. Rather than trying to formulate definitions in a remote viewing context, we will describe what often appears to be unique in public demonstrations of remote viewing and perhaps what was special during the 20-year, US Government program in remote viewing. We remind the reader that these ideas are qualitative and hope that they might inspire quantitative tests of the concepts.

There appears to be something special about conducting remote viewing demonstrations for the media. Certainly the psychosocial conditions, participants' ego, and monetary incentive contribute to the general success. Here, however, we focus on attention, intention, and expectation.

Attention

Of the three concepts in this Section, *attention* is, perhaps, the easiest to discuss within the framework of successful demonstrations and laboratory remote viewing sessions. Part of *attention* is obvious and straight forward; during a trial, the participants should be focused upon the remote viewing task at hand and not be diverted with other simultaneous activity such as eating lunch. However, we suspect that *attention* may be a deeper concept than this.

Qualitatively, the 20+ year US Government program produced results that were consistently successful. If it were actually true quantitatively, then there may be a number of explanations. For example, we are told that observation theory might suggest that those successes arose because of the program's secrecy.[1] However, in our view, other reasons are more likely. To illustrate we rarely used so-called un-selected participants, unlike many laboratories conducting psi research. Rather, we focused upon a few individuals who could reliably produce psi under laboratory conditions and worked with them, in some cases, for as long as 30 years. Perhaps this, alone, could account for the apparent success.

[1] Private communication from Dick Bierman.

Attention, however, is our candidate for a major component for the successes. We illustrate with a number of examples. When we were asked to conduct RV sessions against military or intelligence targets[2] in an application setting, the whole laboratory stopped all activity and focused on the job at hand. At the height of the program this meant that 12 people dropped what they were doing and devoted their efforts in providing the best environment in which to conduct the session. Often such a circumstance included flying one or more viewers to California from across the US. Clearly the *attention* was sharply focused. This same kind of *attention* was afforded to laboratory studies as well.

In contrast, we conducted a laboratory series of 75 trials that showed little evidence of RV (May, Spottiswoode, & Faith, 2000). What was different? This study had two experienced experimenters (one in Palo Alto and the other in Los Angeles) and five experienced viewers scattered all over the US. Rather than flying them all to LFR's remote viewing facility, we designed a complex, internet-based procedure, so for a given trial, each experimenter worked for about an hour and each viewer perhaps only 15 minutes in a telephone interview. Then all participants returned to their daily activity. Surely there may have been intense *attention* during the hour or so, but after the fact we all felt a substantial difference of the kind of *attention* from what was routine in the former government program. For example, a 15-minute scheduled telephone call to a viewer might have interrupted that viewer who was writing a paper, doing taxes, or having lunch. The internal, or perhaps, cognitive rational, might have been, "Oh well, I'll finish what I was doing when I was interrupted when the session is over."

It is difficult and perhaps dangerously misleading to ascribe an apparent psi experiment failure to a qualitative post-hoc discussion like the one above; however, we feel it is important to try to open a dialog on what might be an important, but difficult topic—*attention.* The successful examples were selected from a very large dataset of similar circumstances, but so far we have only the one unsuccessful experiment that could be easily attributed to lack of *attention*.

If one would like to understand the difference between attention in remote viewing trials as expressed within the military project versus a more normal level of attention expressed in an experimental series, we recommend reading the dedication page of *The Stargate Chronicles.*[3]

Intention and Expectation

We include these two concepts in the same Section, because the differences are subtle but important. What we mean by *expectation* is what the whole team desires consciously (and perhaps unconsciously) for the outcome. Fear of psi, or lowered *expectation,* has been part of the lore[4] and literature (Tart, 1984) for some time. If Tart is correct, while we might overtly *expect* a wildly successful trial, unconsciously we might be afraid of it and lower our *expectation* accordingly.

A laboratory anecdote might also illustrate *expectation*. A post doctoral student joined the US Government program late in its history. The other laboratory personnel had high *expectation* for success in whatever we tried based upon nearly two decades of experience. The new post doc was, understandably, skeptical of our success rates, especially when she/he was informed that we had an inverse file drawer problem; that is, we had a number of laboratory studies that were significant and not yet published.[5] The new post doc's *expectation* for success was not as high as that of the other laboratory personnel.

It may seem obvious that high *expectation* is a necessary ingredient for success in any activity, let alone, remote viewing. If you expect to fall on the expert sky run, you probably will. It is not clear to us, however, the degree to which this might impact remote viewing. We have seen high quality remote viewing examples conducted before people who had high expectation of failure.

[2] Almost always in a blind or double-blind protocol.
[3] The Stargate Chronicles, J.W. McMoneagle, Hampton Roads Publishing 2002.
[4] Private communication as part of Ingo Swann's concept of experimenters' anti-psychic syndrome.
[5] That is still true today. LFR, as the keeper of the science record for the US Government Program has a substantial backlog of successful experiments to be published.

Intention while related to *expectation* may be different. We could, for example, gather a team together with the *intention* of winning the lottery but the *expectation* of winning might be rather low. *Intention* when mixed with *expectation* can be used to offer excuses for a potential failure. For example, the spectrum of the florescent lights is wrong; the experimenter did not show up on time; there is too much noise outside, etc. It is possible to overcome these "excuses" by adopting the proper attitude that these external things simply do not matter.

NATIONAL GEOGRAPHIC CHANNEL REMOTE VIEWING TRIAL

What characterizes the 44 public demonstrations in which Joe has contributed is that financially and for reputation reasons the TV production crews and the viewer all had high intention, attention, and expectation. The public displays and filming were conducted more like the days of the US government program than what is done in many laboratory studies. Chuck Honorton first noticed this on-camera effect with the now well-known ganzfeld trial by Ellen Messer with Las Vegas as the target.6

We provide this example as an illustration of combined intention, attention, and expectation. This particular trial was conducted with the best and most secure protocol of any of the media trials so far. In this Section we describe that protocol, in detail, and graphically show the results.

Protocol

The participants for this trial were as follows:

- The show's producer, MH.
- Assistant from the production company, RC.
- Assistant from a local law office, PS.
- Two camera men from Canada, CM1 & CM2
- Additional camera woman from the Bay Area, BA.
- Remote viewer, JM.
- Interviewer and judge, EM.

Thus, a total of eight individuals were actively involved in this single remote viewing. The trial was conducted in the office of the Laboratories for Fundamental Research in Palo Alto, California on Saturday, 21 February 2004. On the previous Thursday and prior to any other of the production crew coming to the Bay Area, RC came to the laboratory directly from London.

Target Preparation Instructions for RC

EM instructed RC on how to choose potential sites for the target, and RC was asked to find six locations within about 25 km from the laboratory and to select them to be as different from one another as possible. Each potential site was to be out of doors and contain elements that could be easily sketched. RC was asked to photograph each site from a number of perspectives and obtain a set of double prints for each.

RC prepared two sets of six opaque envelopes. Each set contained the individual photographs for each of the targets. To differentiate the sets, one was to be marked on a corner of each of the envelopes in the set.

When completed, both sets—one marked and one not—were to be delivered to PS no later than Friday at 17:00 hours. At no time was RC to have verbal or visual contact with any other person involved with the trial until the trial was complete. PS was asked to place one of the two sets aside, shuffle the remaining set; and then number that set from one to six on the outside of the unopened envelopes. At this point, no one knew what photographs were contained in any specific numbered envelope.

[6] If we recall correctly, Chuck gathered together about eight similar cases with seven first place hits.

Trial Protocol

At 10:00 Saturday, JM and EM went to the laboratory to set up for the trail. Shortly after MH and CM1 & CM2 also appeared to set the lighting and camera for the 10:30 start time. MH instructed JM and EM on what to do for the camera. It is important to note that all people in the laboratory were blind to all the possible targets. They did know, however, that all sites were within 25 km of the laboratory.

At 09:30, RC went to PS's office, rolled a 6-sided die to choose one of the six target envelopes, and left with that unopened envelope. Once in the car, RC opened the envelope and drove to the specified target site in order to be there by 10:30.

At 10:30, EM showed JM a photograph of RC and tasked JM. "Have you seen the woman before?" (No was the answer.) "Please describe the location where she is currently standing." For the next 10 minutes JM wrote and drew his impression while on camera. EM asked for clarification and expansion on a number of points, but was silent for most of the session.

At the completion of the session, MH photocopied the response, gave the original to EM, and left the laboratory with JM, CM1 & CM2, and the camera gear. Once in their van, MH called RC to find where to take JM for feedback.

Meanwhile in the laboratory, EM was joined by BA and the laboratory door was closed. BA called PS on the phone and asked that he deliver through the mail slot in the door, the *second* set of six, unnumbered envelopes. Then, on camera, EM performed a rank-order assessment of which of the sets of photographs best matched the response, second best matched and so on for the remaining sets. The result of this ranking was written down for later.

While the blind ranking was in progress, JM, MH and CM1 & CM2 were all located at the intended site for feedback and totally out of communication with anyone back at the laboratory.

When JM, MH, RC and CM1 & CM2 returned to the laboratory, EM provided MH with the results of the blind ranking

Results

EM picked the correct target in first place. By definition the p-value for the single trial was 1/6 = 0.167; however, as we will show below, rank order analysis is a conservative estimate of the information contained in a first-place match. It is beyond the scope of this paper to include the complete 6-page response, but we will include selected drawings and the following verbatim list in JM's final summary page:

- Unique access – tunnel like/passage access under an overpass.
- Building adjacent to what she's handling – work of art.
- Garden. Formal garden – fills a smaller area.
- Paths – both formal – straight lines & informal.
- Central work of art.
- Half arches.
- TGT stands alone – in a clearing w/trees.
- Smooth & rough/metal & stone.
- Large – looks upward at it.

Figure 1 shows one photograph from each of the potential targets.

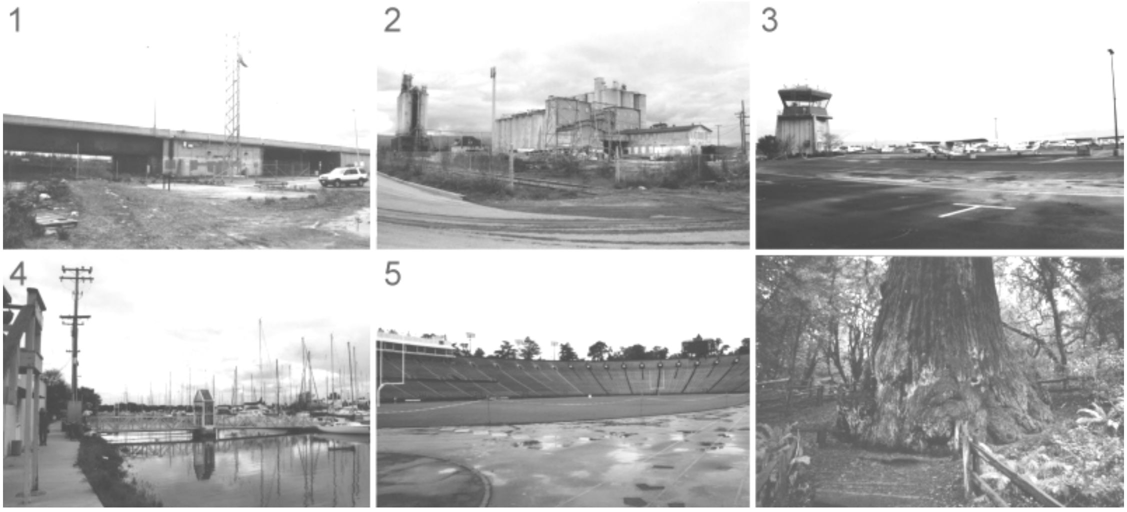


Fig. 1 One Photograph from Each Potential Target Site

The numbers in the upper left corner of each photograph are the rank numbers assigned by EM. EM correctly assigned the end of the Dumbarton Bridge in first place. In order, the other sites were a rock quarry facility, the Palo Alto Municipal air port, the Redwood City marina, the Stanford University football stadium, and a redwood tree in a local park.

Figure 2 shows the correspondence with JM's first impression with regard to access to the site where RC was standing during the trial.

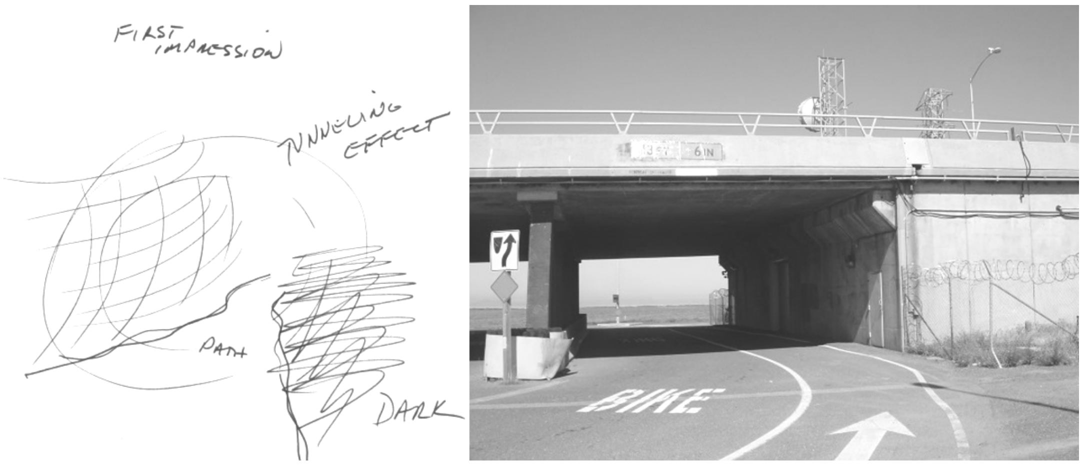


Fig. 2 Tunnel Underpass to the Site

RC was standing immediately on the right just through this tunnel. What was not too clear from the judging photograph (i.e., number 1 in Figure 1 above) was that precise location. RC was standing next to a plaque (i.e., work of art) directly beneath the massive concrete and metal of the ramp to the Dumbarton Bridge. During the session, RC was looking up at the structure, wandering around a "formal" garden and admiring the "work of art." Figure 3 emphasizes these comparisons.

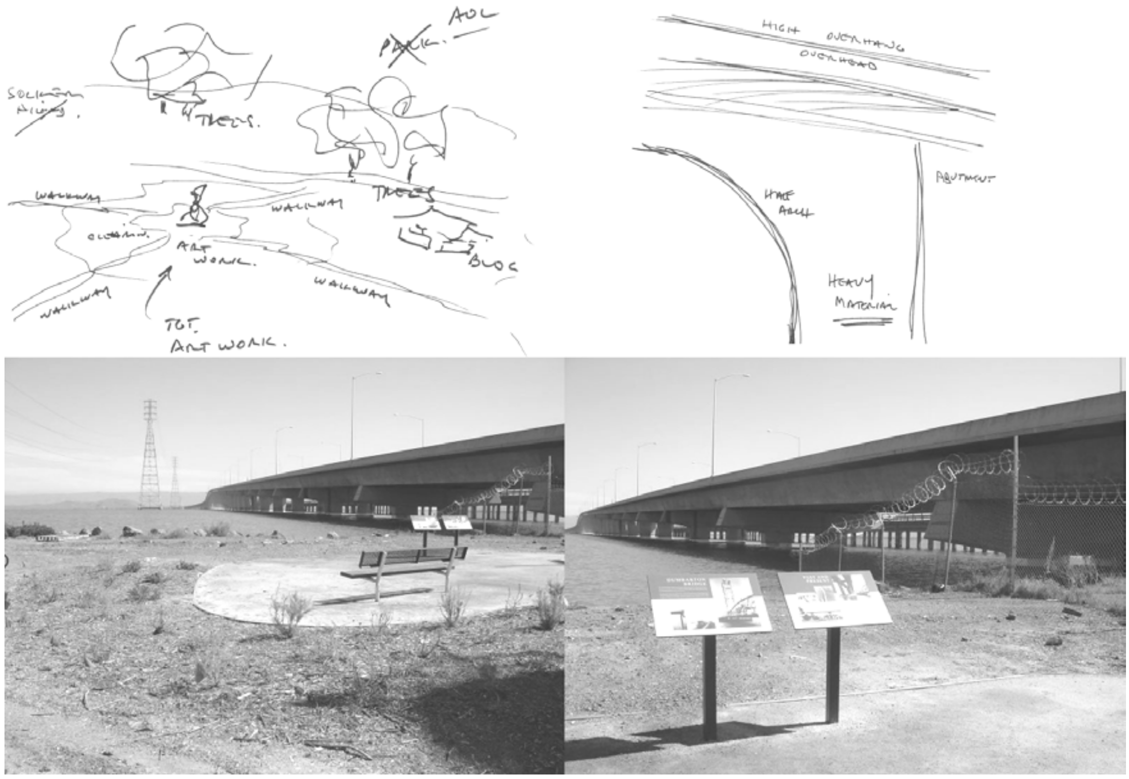


Fig. 3 Garden, Small Area, Work of Art and "High Overhang Overhead"

Unlike the photographs used in the blind rank-order analysis, the ones in Figures 2 & 3 were taken after the fact to correspond to the responses.

With regard to the summary list above, all points were correct with the possible exception of the half arches and a lack of trees and a building. EM experienced difficulty in the ranking between the first and second place using only the a priori photographs; however, if the ranking had been carried out as in the original SRI International protocol by visiting, in turn, each of the sites, there would have been no difficulty, whatsoever, in picking the correct target site in first place.

Finally we comment about rank-order in general. Imagine two extreme cases of remote viewing quality: (1) a near perfect description of the correct target, and (2) barely enough information to allow an analyst to match the response to the correct target. In the rank-order statistic both responses only receive a credit as a 1-in-N match—N = 6 in the example above. Clearly this approach, while conservative, ignores the additional information obtained in the viewing for the first case above. Additionally, using rank-order as a statistic while searching for correlations of remote viewing with other variables (e.g., LST, personality, GMF) may mislead or underestimate the correlation. We spent considerable effort over the last 20 years to develop other methods of analysis that were more sensitive to the information content (May et al., 2000; May, Utts, Humphrey, Luke, & Frivold, 1990).

Discussion

What constitutes a psi-favorable environment is not well understood, but is, nonetheless important. There are a number of cases where it might be assumed that the circumstances would not be psi favorable (e.g., sea-sick viewer, demonstrations that had to be successful to obtain the next contract, next to 155 mm Army gun range, etc) but often high quality remote viewing was obtained even then. What is clear, at least to those of us on our long-established team, is that intention, attention, and, expectation play a very important role in the success of both application and research of remote viewing

ACKNOWLEDGEMENTS

We thank the numerous networks and producers and crews whose intention, attention and expectation contributed to the successes of many of the demonstration remote viewings.

REFERENCES

May, E. C., Spottiswoode, S. J. P., & Faith, L. V. (2000). Correlation of the gradient of Shanon entropy and anomalous cognition: Toward an AC sensory system. *Journal of Scientific Exploration, 14*(1), 53-72.

May, E. C., Utts, J. M., Humphrey, B. S., Luke, W. L. W., & Frivold, T. J. (1990). Advances in remote-viewing analysis. *Journal of Parapsychology, 54*, 193-228.

Tart, C. T. (1984). Acknowledging and dealing with the fear of psi. *Journal of the American Society for Psychical Research 78*(2), 133-143.

Address for correspondence: Edwin C. May, Laboratories for Fundamental Research, 415 Cambridge Ave., Suite 3, Palo Alto, California 94306, USA.
E-mail: may@lfr.org

PHYSIOLOGICAL CORRELATES OF ESP: HEART RATE DIFFERENCES BETWEEN TARGETS AND NON TARGETS IN CLAIRVOYANCE AND PRECOGNITION FORCED CHOICE TASKS

Luisa Sartori, Stefano Massaccesi, Massimiliano Martinelli & Patrizio E. Tressoldi
Department of General Psychology, Padova University

ABSTRACT

Physiological reactions are still considered by most contemporary theories to take place only after stimulation has occurred. Yet recent studies have suggested that the autonomic nervous system can act as a reliable predictor of a future experience (Radin, 1997; Radin & Bierman, 1997, 1998; Bierman, 2000; Radin, 2003; Spottiswoode & May, 2003). Physiological reactions to incoming stimuli can occur without perceptual and cognitive encoding, and these are the very first reactions of the human organism.

This paper reports the results of two different experiments carried out in order to examine such physiological effects and collect further data without the use of arousing materials. We opted for very easy decision-making tasks instead of using violent/erotic pictures in order to avoid ethical problems and to find a scientific paradigm that may also be extended to children.

Spinelli (1980) found out that children up to age eight showed remarkable ESP ability. Parapsychological research with children can be challenging.

More specifically, the present study aims to evaluate heart rate differences during the presentation of targets and non-targets in classical clairvoyance and precognition forced-choice tasks. The gambling-like task consists of a serial presentation of four calm pictures: subjects have to guess which one will be randomly selected as a target. In each trial the target was selected automatically by a pseudo-random algorithm: in the clairvoyance condition, targets were selected before participants did the trials; in the precognition one, targets were determined right after subjects have made their choice.

This procedure was repeated for twenty trials. For each picture presentation, a sample of ten heart rate data was collected. The experiments involved 12 participants who together contributed 240 trials.

The first purpose was to independently replicate and extend the results of previous experiments demonstrating the presentiment effect. The second objective was to examine whether the heart is able to receive information still outside the range of conscious awareness.

Results were significant in both experiments: heart rate associated to targets increased at a statistically significant level compared to non-targets in the clairvoyance condition ($p = .0007$; $d = .054$) and in the precognition one ($p = .03$, two-tailed; $d = .039$), whereas the mean of hits scores in the guessing task was close to the chance level. Results furthermore showed a statistically significant Stouffer value as regards within subjects data: we interpreted these findings in light of research on individual reaction patterns (Lacey & Lacey, 1970).

In conclusion, the present results lend support to the hypothesis that Heart Rate may be a reliable physiological variable to detect ESP cognitive information even if targets overt identification is at chance. The heart appears to play a direct role in the perception of future events.

This area still provides much potential for further research.

INTRODUCTION

Can ESP information be detected physiologically? This is a plausible hypothesis assuming that ESP information differs from non ESP information only in how it is obtained and not in how it is processed. Different authors demonstrated that information not detected at an overt conscious level produces physiologically specific modifications (Dimberg, Thunberg, Elmehed, 2000; Kubota, Sato, Murai, Toichi, Ikeda, Sengoku, 2000; Mayer, Merckelbach, deJong, Leeuw, 1999; Bechara, Tranel, Damasio and Damasio, 1996; Bechara, Damasio, Tranel and Damasio, 1997).

In the field of ESP research the search for a physiological index is not a new interest. Beloff (1974) more than 25 years ago reviewed the studies related to this topic. Recently, Bierman and Radin (2000) reported a replication of their finding that skin conductance level, blood volume and heart rate change according to the emotional category of future pictures. This so called pre-sentiment effect has been replicated by Bierman and Scholte (2002) using fMRI data.

Moreover, further evidences actually support the possibility that similar physiological effects may be observed in tasks without emotional contents. Mc Donough, Don & Warren (2002) detected an EEG activity in the gamma band correlated to a forced choice guessing task confirming previous findings of the same authors (Don, Mc Donough and Warren, 1998) as regards event-related brain potentials.

In this paper we decided to use a task very similar to that one used by Mc Donough et al. (2002), but with an easier technology: we recorded heart rate pulses instead of EEG activity. The purpose was to investigate if this task might produce different results with a clairvoyance or a precognitive condition, and we devised two identical experiments simply changing the moment of the target choice. In the clairvoyance experiment the target was chosen before the presentation of the first picture, whereas in the precognitive experiment it was selected after the subject's choice of the target. The main hypothesis was that heart rate could change according to the categories of pictures, targets vs. non targets. The direction of this difference was not predictable in advance because, at our knowledge, there are not similar evidences in literature.

METHODS

Participants

Twelve voluntary subjects were tested. The proportion of male and females was close to 50%. Mean age was 26,5 years-old (range 24-45). Their performance was reimbursed with € 3.

Procedure

Participants seated in a comfortable chair in a soundproof laboratory. A video monitor was located in front of the participant at eye level and a computer mouse was held in his/her dominant hand. They were instructed not to move their index finger of the nondominant hand, connected to the apparatus detecting their heart rate, to relax when an acoustic signal was perceived and to concentrate on the picture that would be shown for about 10 seconds on the monitor (depending on the time necessary to collect 10 heart rate data based on inter-beat intervals) until it disappeared. This sequence would have been repeated four times with different pictures that subsequently would have been presented simultaneously on the monitor. After the presentation of the four pictures, the participants were invited to guess the target by clicking with the computer mouse the selected picture. After the choice, the real target was illuminated to inform the subjects about the accuracy of his/her choice.

The experiment ended after 20 trials, always using different pictures. Pictures represented coloured calm images, i.e. landscapes, plants, flowers, portraits. Their degree of emotionality was measured asking to ten independent judges to rate each picture on a ten points scale from 0 (no emotion) to 10 (high level of emotion). The mean score was 1.5, SD .5.

The sequence of events for each trial is illustrated in Fig. 1.

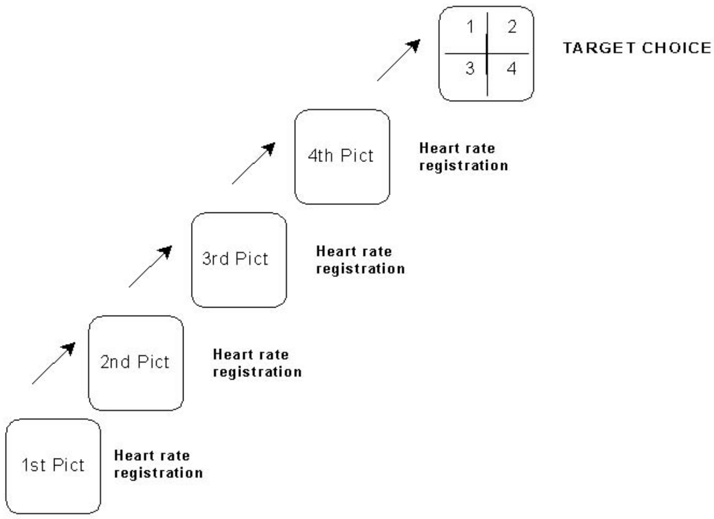


Fig. 1: Scheme of the sequence of events for each trial

The target was selected among the four pictures by an automatic randomisation procedure. The Randomisation procedure written in C++ for these experiments returned a random number within the range 1-4 (corresponding to the four pictures) after an initiation with a random algorithm seeded by computer clock.

Data acquisition and apparatus functioning were continuously monitored by a research assistant placed behind the subject. Owing to the automatic procedure, the research assistant could not have suggested anything. If some artefacts were noticed (for instance anomalous heart rate registrations or apparatus malfunctioning), the task was interrupted and restarted again.

Half subjects performed first the clairvoyance experiment, the other half the precognition one.

Heart rate sampling.

For each picture ten heart rate samples were obtained by a connection between the Cardiofrequencimeter and the computer. The Cardiofrequencimeter consisted of an opto-electronic sensor for a photopletismographic measurement by infrared light applied to the index finger of the nondominant hand. The signal was conveyed to a cardiofrequencimeter Pulse Monitor-701 and to a digital multimeter Metex 3850 D and subsequently fed in a PC for online data acquisition. These data were obtained from the parallel port from which the analogical signal was converted in digital form. The software for pictures presentation and heart rate data acquisition was original and devised for these experiments (Massaccesi, 2001).

Data analysis

The initial 800 data of each subjects (10 seconds x 4 pictures x 20 trials) were reduced to 400 (10 seconds x 20 trials) collapsing the data related to targets and non targets within each trial (200 for targets and 200 for non targets). A simple paired t test was used to compare hear rate means related to target and non targets.

RESULTS

Clairvoyance experiment.

The overall comparison between targets and non targets yielded a statistical significant difference of 0.56, (C.I. 95% ± .33), paired t test: $t_{(2399)} = 3.4$; $p = .0007$; Effect Size $d = .054$ (Dunlap. Cortina, Valow, Burke, 1996). To control the reliability of this result we implemented a bootstrap analysis[1] with the Simstat™ software (Péladeau, Lacoutre, 1993) using 1000 resamples. The result was:

$t_{(2399)} = 3.38$ (C.I. 95 ± 1.98).

The mean of correct hits was close to the chance level, M = 6.08; SD = 1.37.

Precognition experiment

The overall heart rate raw difference between target and non target of 0.39 (C.I. 95 ± .36) was statistically significant, $t_{(2399)} = 2.16$; $p = .03$ (2-tailed); Effect size $d = .039$. This result was confirmed using the bootstrap procedure using 1000 resamples, $t_{(2399)} = 2.19$ (C.I. 95% ± 2.03).

As in the clairvoyance experiment, the mean of hits was at chance level, M = 5.6, SD = 2.05.

Experiment comparison

The differences between targets and non targets in the two experiments are better illustrated in Figure 2a and Figure 2b.

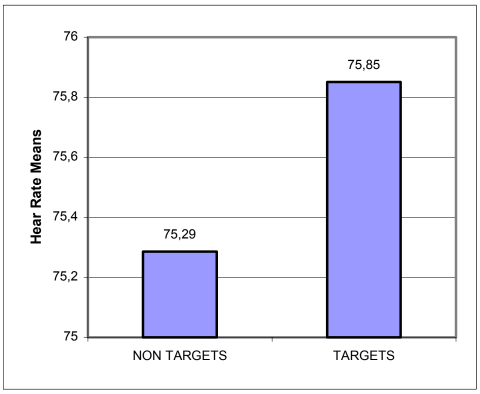


Fig. 2a Heart rate means of targets and non targets in the clairvoyance experiment.

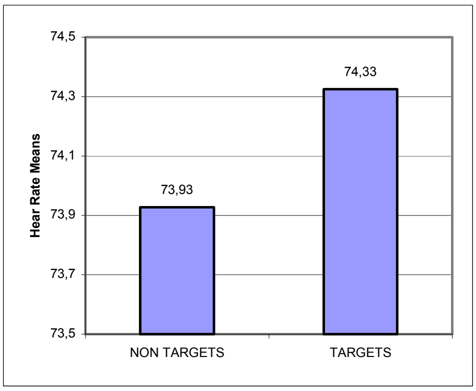


Fig. 2b Heart rate means of targets and non targets in the precognition experiment.

The differences between targets and non targets in the two experiments are not statistically different $t(2399) = .68$.

[1] Bootstrap simulation is a resampling technique whereby initial sample are treated as if they constitute the population under study. By replicating those data an infinite number of times, we then draw at random from that population a large number of samples, each one the same size as the original sample. By computing, for every bootstrap sample, a statistical estimator of interest (such as a mean, a correlation or a t-test between two variables), this resampling procedure recreates an empirical sampling distribution of this estimator. The main advantage of such a procedure is that the sampling distribution is not mathematically estimated but empirically reconstructed with all the original characteristics of the data. So, it takes automatically into account distribution properties that are generally considered as contaminating factors, such as skewness, ceiling effects, outliers, etc. This feature makes bootstrap estimations adequate even when data are not normally distributed. In fact, bootstrapping can even be used to describe the sampling distribution of estimators for which sampling properties are unknown or unavailable.

Meta analysis using single data.

Given the high number of data for each subject it is possible to analyse the reliability of the differences between targets and non targets in the two experiments calculating a Stouffer z from the t values obtained from each subject. The individual t test and the Stouffer value are presented in the Table 1.

Table 1: Individual t values of each subject in the two experiments with Stouffer z value.

	Clairvoyance	Precognition
X1	-0,99	-0,55
X2	-1,48	-0,19
X3	-1,42	0,26
X4	0,61	-1,17
X5	-0,88	0,7
X6	-3,1	-1,48
X7	-1,39	-0,2
X8	-1,49	0,78
X9	2,43	-3,4
X10	-1,63	-0,05
X11	-2,57	-2,64
X12	0,06	1,7
Stouffer z	-3,42	-1,8
p (2-tailed)	0,0003	0,036

For both experiments the Stouffer value is statistically significant and also their difference,
$z = 1.6$; $p = .054$.

DISCUSSION

With a simple procedure we have obtained a clear evidence that the heart rate is a physiological variable sensible enough to differentiate two categories of information that will be known in the future. Even if the raw difference is very low, less than one heart rate bit per second on average, it seems quite reliable as demonstrated by the analysis of our data. The difference between the clairvoyance and the precognition experiment seems not very reliable, even if in the second condition the heart rate difference between targets and non targets appears less evident. However, in both experiments, at the overt cognitive level, the means of hits is close to chance.

Our findings offer new evidence that it is possible to detect ESP signals at a neuro-physiological level adding convergent support to the EEG pre-knowledge paradigm investigated by McDonough and coll. (2002). A critique to these "ESP physiological signals" is that not only they are so weak to be practically undetectable by the subjects but also that they are extracted from a relatively high number of trials giving a statistical high power to detect very low effect sizes. This critique is correct, however if these physiological signals are real pre-(overt) cognitive information, we can start to investigate if it is possible to extract a "prototypical pattern" distinguishing targets from non targets at the level of single or less numerous trials; for example, using algorithms implemented in neural networks. Another possibility is to manipulate the "physiological signal" enhancing the differences at a level detectable by the subjects. The theoretical contribution of Stevens (2000) about Stochastic resonance (SR) a phenomenon wherein some characteristics of the signal (amplitude, signal-to-noise ratio, coherence, etc.) are, counter to intuition, actually improved by the presence of the noise, may be another useful approach for this line of investigation.

References

Bechara, A., Tranel, D., Damasio, H., & Damasio, A. R. (1996). Failure to respond to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, 6, 2, 215-225.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293-1295.

Beloff, J. (1974) ESP: the search for a physiological index. *Journal of the Society for Psychical Research*, 47, 403-420.

Bierman, D. & Radin, D.I. (2000) Anomalous unconscious emotional responses: evidence for a reversal of time. In: S.R.Hameroff, A.W.Kaszniak & R.D.Chalmers (Eds) Tucson III: *Towards a Science of Consciousness* (pp. 367-386). MIT Press, Cambridge.

Bierman, D.J. & Scholte, S. H. (2002) Anomalous anticipatory brain activation preceding exposure of emotional and neutral pictures. *Towards a Science of Consciousness*, Tucson, (pp.140).

Dimberg, U., Thunberg, M. & Elmehed, K. (2000) Unconscious facial reactions to emotional facial expressions. *Psychological-Science*. 11,1, 86-89.

Don, N.S., Mc Donough, B.E. & Warren, C.A. (1998) Event-related brain potentials (ERP) indicators of unconscious PSI: a replication using subjects unselected for PSI. *The Journal of Parapsychology*, 62,127-145.

Dunlap, W.P., Cortina, J.M., Valow, J.B. & Burke, M.J.(1996) Meta-analysis of experimental with matched groups or repeated measures designs. *Psychological Methods*, 1,2: 170-177.

Kubota, Y., Sato, W., Murai, T., Toichi, M., Ikeda, A. & Sengoku, A. (2000) Emotional cognition without awareness after unilateral temporal lobectomy in humans, *Journal of Neuroscience*, Oct; 20,19, RC97.

Massaccesi, S. (2001). Un sistema modulare portatile per la rilevazione dell'attività elettrodermica e della frequenza cardiaca (a portable modular system for electrodermic and heart rate acquisition). *Giornale Italiano di Psicologia*, 4, 825 – 837.

Mayer, B., Merckelbach, H., deJong, P. & Leeuw, I. (1999) Skin conductance responses of spider phobics to backwardly masked phobic cues. *Journal of Psychophysiology*, 13, 3,152-159.

Mc Donough, B.E., Don, N.S. & Warren, C.A. (2002) Differential event-related potentials in targets and decoy in guessing task. *Journal of Scientific Exploration*, 16, 2, 187-206.

Péladeau, N. & Lacoutre, Y. (1993) SIMSTAT: bootstrap computer simulation and statistical program for IBM personal computer. *Behavioral Research Methods Instruments & Computer*, 25,3, 410-413.

Stevens, P. (2000) Noise, physics and psi: new ideas for research. *International Journal of Parapsychology*, 11,63-72.

Address for correspondence: Patrizio E. Tressoldi,
Dipartimento di Psicologia Generale, Via Venezia, 8,
35131 Padova, Italy.
E-mail: patrizio.tressoldi@unipd.it

A PRIMARY QUANTUM MODEL OF TELEPATHY

Gao Shan
The Scientists Work Team of Electro-Magnetic Wave Velocity,
Chinese Institute of Electronics and Institute of Quantum Physics, Beijing

ABSTRACT

The physical nature of psi phenomena such as telepathy is an important problem in present science of consciousness. Scientists have basically confirmed the existence of telepathy phenomena through many strict experiments. Then can modern science (e.g. quantum theory) provide a scientific explanation for telepathy phenomena? In this paper, we will seek the possible quantum nature of telepathy from both theoretical and experimental aspects, and will present a primary quantum model of telepathy process. It is well known that even though present quantum theory permits the existence of quantum nonlocality, it does forbid the realization of nonlocal communication or quantum superluminal communication (QSC). However, the usual no-go theorems don't consider the possible active role of consciousness during quantum measurement process. In a recent paper (see Found. Phys. Lett, 17(2), 167-182), it has been demonstrated that a proper combination of quantum process and conscious perception will permit the distinguishability of nonorthogonal quantum states, and further result in the realization of QSC or nonlocal communication. This is called the QSC principle. On the basis of the QSC principle, we propose a primary theoretical model of telepathy process. According to the model, the telepathy process mainly includes three phases. The first phase is to form the quantum entanglement state of brains, the second phase is to hold the entanglement state of brains, and the third phase is to collapse the entanglement state of brains. When the entanglement state of brains is collapsed by a certain measurement on one of the subjects, the brain states of both subjects turn to be definite states from entanglement state, and the other subject will perceive the change at a distance according to the QSC principle. When in the entanglement state or superposition state, no definite perception relating to the state exists, whereas when the superposition state collapses into a definite state, a definite perception relating to the collapse state appears. Then the telepathy between the subjects may appear. It should be stressed that, even though the above quantum model may in principle provide a primary scientific explanation of telepathy phenomena, there are still some left technical problems such as the expression of high-level telepathy information etc. Lastly, in order to test the QSC principle and the above quantum model of telepathy, some feasible schemes of quantum perception experiments and perception entanglement experiments are further proposed on the basis of present technology. We urge that such quantum perception experiments need to be conducted as soon as possible. If the experiment results are positive, they will have far-reaching influence on the present science of consciousness and psi research, and will help to bridge the gap between the parapsychology and present science.

INTRODUCTION

The physical nature of psi phenomena such as telepathy is an important problem in present science of consciousness. The existence of telepathy phenomena has been basically confirmed through many strict experiments (Duane & Behrendt, 1965; Targ & Puthoff, 1974; Puthoff & Targ, 1976; Radin & Nelson, 1989; Grinberg-Zylberbaum et al, 1994; Bierman & Radin, 1997; Gao, 2000; Wackermann et al, 2003). Then can modern science (e.g. quantum theory) provide a scientific explanation for telepathy phenomena? In this paper, we will seek the possible quantum nature of telepathy from both theoretical and experimental aspects, and will present a primary quantum model of telepathy phenomena. It will be shown that, according to the principle of quantum superluminal communication (QSC) (Gao, 2000; Gao, 2003; Gao, 2004), quantum theory can in principle provide a scientific explanation of telepathy phenomena when considering the role of consciousness in quantum process, and some experiments may have indicated the validity of this explanation. Lastly, we will propose a series of feasible experiment schemes to test the quantum model of telepathy.

A ROLE OF CONSCIOUSNESS IN QUANTUM MEASUREMENT

We will first analyze the role of consciousness in physical measurement process. As we know, physical measurement generally consists of two processes: (1). the physical interaction between the observed object and measuring device; (2). the psycho-physical interaction between the measuring device and observer. In some special situations, measurement may be the direct interaction between the observed object and observer.

Even though what physics studies is the insensible object or matter, the consciousness of the observer must take part in the last phase of measurement. The observer is introspectively aware of his perception about the measurement results. The conscious function is used to end the infinite chains of measurement. This is the main role of consciousness different from that of usual measuring device in the measurement process. But this difference doesn't result in the physically testable different displays for classical measurement process. In classical theory, the influence of the measuring device or observer to the observed object can be omitted in principle during measurement process, and the psycho-physical interaction between the observer and measuring device does not influence the reading of the pointer of the measuring device either. Thus classical measurement is only one kind of plain one-to-one mapping from the state of the observed object to the pointer state of the measuring device and further to the mental state of the observer, or direct one-to-one mapping from the state of the observed object to the mental state of the observer from a physical point of views. In short, the consciousness of the observer possesses no physically testable different functions from physical measuring device in classical theory.

However, the measurement process is no longer plain in quantum theory. The influence of the measuring device to the observed object can't be omitted in principle during quantum measurement owing to the existence of quantum superposition. It is just this influence that generates the definite measurement result to some extent. Since the measuring device has generated one definite measurement result, the psycho-physical interaction between the observer and measuring device is still one kind of plain one-to-one mapping, and this process is the same as that in classical situation. But when the observed object and observer directly interact, the existence of quantum superposition will introduce new element to the psycho-physical interaction between the observer and measured object. The interaction will result in the appearance of the observer with consciousness in quantum superposition state. Then whether or not does the consciousness of the observer in quantum superposition state have some physically testable different displays from physical measuring device? We will try to give the answer.

In order to further analyze the possible role of consciousness during quantum measurement, we need a complete theory describing the quantum measurement process. As we know, present quantum theory hasn't provided a complete description of the evolution of wave function during measurement yet, and the projection postulate is just a makeshift. Revised quantum dynamics (Ghirardi et al, 1986; Pearle, 1989; Diosi, 1989; Ghirardi et al, 1990; Penrose, 1996; Gao, 2000; Gao, 2001; Gao, 2003) and many-worlds theory (Everett, 1957; Dewitt et al, 1973; Deutsch, 1985) are two main alternatives to a complete quantum theory. Here we mainly discuss the measurement process in the framework of revised quantum dynamics, and the conclusion will be also valid in the many-worlds theory. At the present time, even if the last complete theory has not been found, but one thing is certain, i.e. the collapse process of wave function is one kind of dynamical process, and it will take a finite time interval to finish. Our analysis will only rely on this common character of the complete quantum theory.

As we know, the nonorthogonal quantum states or nonorthogonal states such as ψ_1 and $\psi_1 + \psi_2$ can't be distinguished in present quantum theory. What's more, the usual measurement using physical measuring device can't distinguish the nonorthogonal states either in the framework of revised quantum dynamics. But when the physical measuring device is replaced by a conscious being and considering the influence of consciousness, it has been shown that the nonorthogonal states can be distinguished under some condition using the consciousness function (Gao, 2000; Gao, 2003; Gao, 2004). The observer with consciousness may obtain more information about the intermediate process before the collapse of wave function finishes, and the added information can help him distinguish the nonorthogonal states. This is the special role of consciousness different from that of physical measuring device during quantum measurement process. Here we will introduce the main ideas.

Let the states to be distinguished be the nonorthogonal states ψ_1 and $\psi_1+\psi_2$, where ψ_1 and ψ_2 can trigger the definite perception states χ_1 and χ_2 of the observer, and the initial perception state of the observer be χ_0. After interaction the corresponding entangled state of the whole system is respectively $\psi_1\ \chi_1$ and $\psi_1\ \chi_1+\psi_2\ \chi_2$. We assume that the observer satisfies the "QSC condition", i.e., his perception time for the definite state $\psi_1\ \chi_1$, which is denoted by t_P, is shorter than the dynamical collapse time of the superposition state $\psi_1\ \chi_1+\psi_2\ \chi_2$, which is denoted by t_C, and that the time difference $\Delta t = t_C - t_P$ is large enough for him to identify. The observer can perceive the input definite state ψ_1 after the perception time t_P, whereas for the input superposition state $\psi_1+\psi_2$, only after the collapse time t_C can the observer perceive the collapse state ψ_1 or ψ_2. Before the collapse time t_C the observer in superposition state $\psi_1\ \chi_1+\psi_2\ \chi_2$ has no definite perception related to the definite state ψ_1 or ψ_2; After the collapse time t_C, the state of the measured system collapses to a definite state ψ_1 or ψ_2, and the observer has a definite perception for the collapse state ψ_1 or ψ_2. Since the observer can be conscious of the time difference between t_P and t_C, he can distinguish the nonorthogonal states ψ_1 and $\psi_1+\psi_2$. It should be stressed that, since the collapse time of a single superposition state is an essentially stochastic variable, which average value is t_C, the "QSC condition" can be in principle satisfied in some collapse processes. For these stochastic collapse processes, the collapse time of the single superposition state is much longer than the (average) collapse time t_C and the perception time t_P.

A NONLOCAL COMMUNICATION METHOD

It is well known that even though present quantum theory permits the existence of quantum nonlocality (Einstein et al, 1935; Bell, 1964; Aspect, 1982), it doesn't permit the realization of quantum superluminal communication (QSC) (Eberhard, 1978; Ghirardi, 1980). However, such demonstrations didn't consider the possible active role of consciousness during quantum measurement process. As we have demonstrated above, a proper combination of quantum process and conscious perception will permit the distinguishability of nonorthogonal states. Once the nonorthogonal states can be distinguished, we can directly realize QSC or nonlocal communication. This is called the QSC principle. Here we present a typical method to realize nonlocal communication.

According to the above analysis, we can design a device to distinguish the nonorthogonal states. We call it NSDD (Nonorthogonal States Distinguishing Device). The design rules are as follows, i.e., when the input state is a definite state, the output of NSDD is '1', whereas when the input state is a superposition of definite states, the output of NSDD is '0'. In the following, we will briefly introduce how to achieve QSC using NSDD.

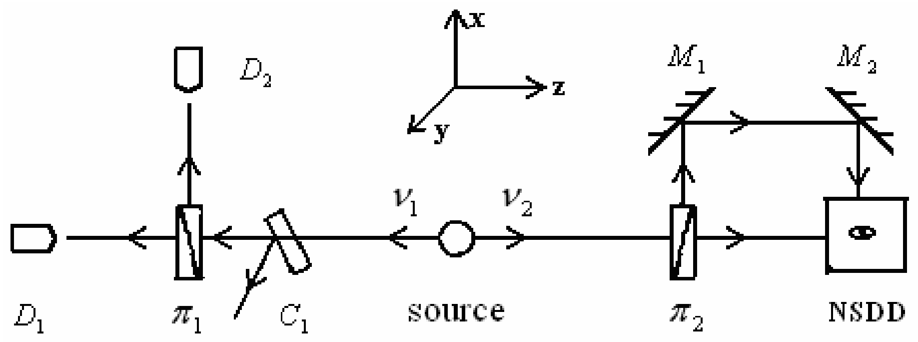


Fig. 1: A scheme of QSC

We use the EPR polarization correlation pairs of photons as the carriers of information. We encode the outgoing information by operating the polarizer, and decode the incoming information using NSDD. The experimental setting is shown in Fig 2. Pairs of photons, whose frequencies are ν_1 and ν_2, are emitted in the -z direction and +z direction from a source, are then analyzed by the two-channel polarizer π_1 and π_2 respectively. The optical switch C_1 in the left side can be controlled to determine whether or not the photon ν_1 will pass to π_1. The transmission axes of the polarizers are both set in the direction x. The two-channel polarizer π_1 and π_2 allows the polarization components of the photon both parallel to and perpendicular to the transmission axis of the polarizer to be passed. The photon passed and analyzed by the polarizer π_1 is detected by D_1 or D_2, and the photon analyzed by the two-channel polarizers π_2 is divided into two paths in space, and respectively input to NSDD from different directions.

We now explain how QSC can be achieved by means of the above setting. Let the sender operate the optical switch C_1, and have the receiver observe the output of NSDD. Suppose the communication rules are stated as follows. The encoding rule for the sender is that not measuring the photon represents sending the code '0', and measuring the photon represents sending the code '1'. The decoding rule for the receiver is that the output of NSDD being '0' represents having received the code '0', and the output of NSDD being '1' represents having received the code '1'. The communication process can be stated as follows. When the sender wants to send a code '0', he controls the optical switch C_1 to let the photon ν_1 move freely and not be analyzed by the polarizer π_1. Then the state of the photon ν_2 is a superposition state after it passes the polarizer π_2, and the output of NSDD is '0'. The receiver can decode the sent code as '0'. When the sender wants to send a code '1', he controls the optical switch C_1 to allow the photon ν_1 to be analyzed by the polarizer π_1 and detected by D_1 or D_2 before the photon ν_2 arrives at NSDD. Then the state of the photon ν_2 collapses into a definite state, and the output of NSDD is '1'. The receiver can decode the sent code as '1'. Thus the sender and receiver can achieve QSC using the above setting and communication rules.

TELEPATHY AND ITS POSSIBLE QUANTUM EXPLANATION

Even though some superphysical phenomena may be not real, telepathy does exist. Its usual display is that one can perceive the other's happening, say being sick or being injured etc, at a distance between the familiar people, say twins, relatives or friends. Many people have this kind of experience. At present, the telepathy phenomena have been basically confirmed by some strict experiments (Duane & Behrendt, 1965;

Targ & Puthoff, 1974; Puthoff & Targ, 1976; Radin & Nelson, 1989; Grinberg-Zylberbaum et al, 1994; Bierman & Radin, 1997; Gao, 2000; Wackermann et al, 2003), and are being studied by more scientists.

In the experiment conducted by Grinberg-Zylberbaum et al (Grinberg-Zylberbaum et al, 1994), pairs of subjects were first allowed to meditate together, and then put into two semisilent Faraday chambers 14.5m apart. Their EEG activities are registered by two EEG machines. One subject of each pair was stimulated by 100 flashes at random intervals, and each photostimulation resulted in an evoked potential for the stimulated subject. It is observed that, when the stimulated subject showed distinct evoked potentials, the nonstimulated subject showed "transferred potentials" similar to the evoked potentials in the stimulated subject, at the same time, the subjects both felt their interaction had been successfully completed. In another experiment conducted by Wackermann et al (Wackermann et al, 2003), six channels electroencephalogram (EEG) were recorded simultaneously from pairs of separated human subjects in two acoustically and electromagnetically shielded rooms. Even though the "transferred potentials" is not found in the experiment, the correlations between brain electrical activities of two spatially separated human subjects are also observed using a more complex method of data analysis. Since the subjects were separated by the soundproof Faraday chambers in the above experiments, these experiments guarantee that neither sensory signals nor electromagnetic signals is the means of communication, and thus strictly confirms the existence of nonlocal correlations and nonlocal communication between human brains.

In the following, we will analyze the above telepathy experiments in terms of the above QSC principle. According to the QSC principle, the proper combination of quantum collapse and conscious perception will result in the realization of nonlocal communication. It will be shown that this may provide a possible explanation of the above telepathy experiments, and indicates that telepathy may result from the quantum process in brains.

We first argue that the "QSC condition" is satisfied in the above telepathy experiment as implied by the experiment results. The "QSC condition" is that the perception time of a conscious being for the definite state is shorter than the dynamical collapse time of the perceived quantum superposition state, and the time difference is large enough for the conscious being to identify. On the one hand, the quantum entanglement state between the subjects A and B in the experiment, which is formed by meditative interaction or other means, can hold for a long time until the experiment is completed, then there appears the observed correlations between brain electrical activities of the two subjects. This indicates that the dynamical collapse time of the quantum entanglement state is also very long, say several ten minutes. On the other hand, the perception time of the subjects for the definite state is generally of the orders of 500ms. Thus in the above experiments the collapse time of the quantum entanglement state or quantum superposition state is evidently much longer than the perception time of the subject for the definite state, and the time difference is also large enough for the subject to identify, i.e., the "QSC condition" is satisfied in the experiments.

It seems to be a well-known fact that the wet and warm brain doesn't support the quantum coherence (Tegmark, 2000). However, on the one hand, the "QSC condition" is related to the collapse time, not the decoherence time of wave function. Even though the decoherence time is very short due to environmental decoherence, the collapse time may be much longer (Hagan et al, 2002). Here we will also give an example. As we know, the number of neurons which can form a definite conscious perception is in the levels of 10^4. In each neuron, the main difference of activation state and resting state lies in the motion of 10^6 Na^+s passing through the membrane. Since the membrane potential is in the levels of 10^{-2} V, the energy difference between activation state and resting state is approximately 10^4 eV. According to one kind of revised quantum dynamical theory (Percival, 1994; Hughston, 1996; Fivel, 1997; Gao, 2000; Gao, 2003), the collapse time of the superposition of the activation state and resting state of one neuron is $\tau_c \approx \frac{\hbar E_p}{(\Delta E)^2} \approx (\frac{2.8Mev}{0.01MeV})^2 \approx 10^5 s$, where $\hbar$ is Planck constant divided by 2π, E_p is Planck energy, ΔE is the energy difference of the state. Thus the collapse time of the superposition of two different conscious perceptions is $\tau_c \approx (\frac{2.8Mev}{100MeV})^2 \approx 1ms$, in which one conscious perception state contains 10^4 neurons in the activation state, and the other conscious perception state contains 10^4 neurons in the resting state. Since the collapse process is an essentially stochastic process, and the collapse time of a single superposition state is a stochastic variable, which average value τ_c is nearly $1ms$, the "QSC condition" can be in principle satisfied in some collapse processes happening in the brains. For these stochastic collapse processes, the collapse time of the single superposition state is much longer than the average perception time $500ms$. This may account for the experimental results of the above telepathy experiments.

Once the required "QSC condition" is satisfied, realizing QSC and explaining telepathy will be probable. According to the QSC principle, the subject satisfying the "QSC condition" will possess different perceptions for the definite state and the superposition of definite states. As revealed in the experiment, when the subject A is not stimulated and the quantum entanglement state still holds, the subject B will be in a superposition state, and he has no distinct feeling or distinct distributions of the brain electrical activities related to the state. Whereas when the subject A is stimulated and the quantum entanglement state collapses, the subject B will be in a definite state, and he does have a distinct feeling that their interaction has been successfully completed or distinct distributions of the brain electrical activities. Then QSC can be realized if we encode the different stimulating operations to subject A, and correspondingly decode the codes through the different feelings or EEG activities of subject B. This may naturally explain the telepathy phenomenon between the subjects.

A QUANTUM THEORETICAL MODEL OF TELEPATHY PROCESS

On the basis of the QSC principle and the above analyses, we will further present a primary theoretical model of telepathy process. In this model, the telepathy process includes three main phases.

Phase 1: Form the quantum entanglement state of brains

During this phase, the quantum states of the brains of the telepathy subjects are entangled. Here we give a possible way to entangle the quantum states of brains. Suppose two photons are in the entanglement state $\psi_1\varphi_2 + \varphi_1\psi_2$, and they respectively enter the eyes of two subjects A and B whose initial states is respectively $\chi_0(A)$ and $\chi_0(B)$. Then after interaction the entanglement state of these two brains will be formed, which can be written as $\chi_1(A)\chi_2(B) + \chi_2(A)\chi_1(B)$. Here we assume that the photons are absorbed in the process. In the above experiments, this phase is achieved by the meditative interaction or other interactions between the subjects.

Phase 2: Hold the entanglement state of brains

The formed entanglement state of brains may hold for a long time in some places of brain under some special conditions. According to the QSC principle, the holding time should be much longer than the usual perception time of the subjects. It is argued that this condition may be satisfied in some places of the brain (Penrose, 1994; Hameroff & Penrose, 1996; Hagan et al, 2002). In the above experiments, the entanglement state is hold by the subjects feeling each other's presence at a distance.

Phase 3: Collapse the entanglement state of brains

When the entanglement state of brains is collapsed by a certain measurement on one of the subjects, the brain states of both subjects turn to be definite states from entanglement state, and the other subject will perceive the change at a distance according to the QSC principle. Here the telepathy between the subjects appears. When in the entanglement state or superposition state such as $\chi_1(A)\chi_2(B) + \chi_2(A)\chi_1(B)$, no definite perception relating to the state exists, whereas when the superposition state collapses into a definite state $\chi_1(A)\chi_2(B)$ or $\chi_2(A)\chi_1(B)$, a definite perception relating to the collapse state can appear. In the above experiment, this phase is achieved by stimulating the subject A with flashes or visual patterns, and when the entanglement state is collapsed by the stimulation, the subjects will display distinct distributions of the brain electrical activities or even feel that their interaction has been successfully completed.

It should be stressed that, even though the above primary quantum model may in principle provide a scientific explanation of telepathy phenomena, there are still some left technical problems. One is to find the position in the brain where the holding time of a quantum superposition state can be much longer than the usual perception time, i.e., to test the existence of "QSC condition" in human brains. Another problem is to study how the brain generates the high-level telepathy information from the low-level one transmitted through the above QSC means. This closely relates to present neuroscience research. Undoubtedly, these unsolved problems need to be deeply studied in experiments. In the next section, we will suggest some experimental schemes that may help to solve the problems.

SOME EXPERIMENTAL SCHEMES

In order to test the existence of "QSC condition" in human brains, and confirm the above primary quantum model of telepathy, we propose the following experimental schemes.

Control experiment

Produce some photons with a certain frequency. Input them to the eyes of the subject. Test and record the conscious time of the subject through EEG (electroencephalograph) or his oral description.

Quantum perception experiment I

Produce the direction superposition state of the photons with the same frequency (e.g. as stated in section 2). Input one branch of the superposition state to the eyes of the subject, and let the other branch freely spread (not input to a measuring device). Test whether the subject perceives the photons during the normal conscious time.

Quantum perception experiment II

Produce the direction superposition state of the photons with the same frequency. Input both branches of the superposition state to the eyes of the subject. Test whether the subject perceives the photons during the normal conscious time.

Perceptions entanglement experiment I

Produce the direction superposition state of the photons with the same frequency. Input the branches of the superposition state to the eyes of two independent subjects respectively. Test whether the subjects perceive the photons during the normal conscious time. It is suggested that the subjects are unfamiliar with each other before the experiment, which can be further confirmed by the phase incoherence of their brain waves.

If the subjects can only perceive the photons after a time interval longer than their normal conscious time in any case of the above experiments, then we will have confirmed the existence of "QSC condition" in human brains. Besides, we suggest that the subjects in the above experiments should be composed of three independent groups at least. The subjects in the first group are in normal state, the subjects in the second group are in meditation state, and the subjects in the third group are in qigong state.

Perceptions entanglement experiment II

Produce the direction superposition state of the photons with the same frequency. Input the branches of the superposition state to the eyes of two independent and isolated subjects respectively. Then stimulate one of the subjects using flashes at random intervals. Record his evoked potentials and the corresponding brain electrical activities of the other subject. Test whether there exists statistical relevance between these brain electrical activities. At the same time, ask the subjects whether they have some kind of conscious perception relating to the stimulations. The existence of the correlation of the brain electrical activities or the direct perception will have confirmed the above primary quantum model, and it can be used to realize controllable human brain communication.

This experiment can be taken as the quantum version of the above telepathy experiments. It is further suggested that the experiment be conducted at much longer distance, e.g. at a distance longer than the bound distance 40km. The present experimental results have shown that the maximum time delay between the EEG response of the receiver and the evoked potentials of the sender is approximately 130 μs (Grinberg-Zylberbaum et al, 1994; Wackermann et al, 2003). Considering the value of light speed, the bound distance excluding the influence of classical signals with light speed is approximately 40km. Here the possible classical signals with light speed can't be used to explain the statistical relevance between the potentials of the subjects. Thus we can strictly confirm that the possible human brain communication is one kind of superluminal and non-electromagnetic phenomena, and further confirm the proposed quantum model of telepathy.

References

Aspect A., Dalibard J. & Roger G. (1982). Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Phys.Rev.Lett 49*, 1804.

Bell J.S. (1964). On the Einstein Podolsky Rosen Paradox. *Physics 1*, 195.

Bierman, D.J. & Radin, D. I. (1997). Anomalous anticipatory response on randomized future conditions. *Perceptual and Motor Skills 84*, 689-690.

Deutsch, D. (1985). Quantum Theory as a Universal Physical Theory. *Int. J. Theo. Phys. 24*, 1-41.

DeWitt, B. S. & Graham, N. ed. (1973). *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton: Princeton University Press.

Diosi, L. (1989). Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev. A, 40*, 1165-1174.

Duane, D. & Behrendt, T. (1965). Extrasensory electroencephalographic induction between identical twins. *Science, 150*, 367.

Eberhard P .H. (1978). Bell's theorem and the different concepts of locality. *Nuovo Cimento B, 46*, 392.

Einstein A., Podolsky B. & Rosen N. (1935). Can quantum mechanical description of physical reality be considered complete? *Phys. Rev. 47*, 777-780.

Everett, H. (1957). "Relative state" formulation of quantum mechanics. *Rev. Mod. Phys. 29*, 454-462.

Fivel, D. I. (1997). An indication from the magnitude of CP violations that gravitation is a possible cause of wave-function collapse, LANL e-print quant-ph/9710042.

Gao Shan (2000). *Quantum Motion and Superluminal Communication.* Beijing: Chinese BT Publishing House.

Gao Shan (2001). From quantum motion to classical motion-seeking the lost reality. *Physics Essays*, *14(1)*, 37-48.

Gao Shan (2003). *Quantum.* Beijing: Tsinghua University Press.

Gao Shan (2004). Quantum collapse, consciousness and superluminal communication. *Found. Phys. Lett*, *17(2)*, 167-182.

Ghirardi G.C., Rimini A. & Weber T. (1980). A general argument against superluminal transmission through the quantum mechanical measurement process. *Letters Nuovo Cimento. 27*, 293.

Ghirardi,G.C., Rimini, A. & Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D*, *34*, 470-491.

Ghirardi,G.C., Rimini, A. & Weber, T. (1990). A Continuous-spontaneous-reduction model involving gravity. *Phys. Rev. D*, *42*, 1057-1064.

Grinberg-Zylberbaum, J., Dalaflor, D., Attie, L. & Goswami, A. (1994). The Einstein- Podolsky-Rosen paradox in the brain: The transferred potential. *Physics Essays*, *7*, 422.

Hagan S., Hameroff S. R. & Tuszynski J. A. (2002). Quantum computation in brain microtubules: Decoherence and biological feasibility. *Phys. Rev. D, 65*, 061901.

Hameroff, S. R. & Penrose, R. (1996). Conscious events as orchestrated space-time selections. *Journal of Consciousness Studies*, *3 (1),* 36-53.

Hughston, L. P. (1996). Geometry of Stochastic State Vector Reduction. *Proc. Roy. Soc. Lond. A 452*, 953-979.

Pearle, P. (1989). Combining stochastic dynamical state-vector reduction with spontaneous localization. *Phys. Rev. A 39*, 2277- 2289.

Penrose, R. (1996). On gravity's role in quantum state reduction. *Gen. Rel. and Grav*, *28*, 581-600.

Penrose, R. (1994). *Shadows of the Mind: a search for the missing science of consciousness.* Oxford : Oxford University Press.

Percival, I. C. (1994). Primary state diffusion. *Proc. Roy. Soc. Lond. A 447*, 189-209.

Puthoff, H. E. & Targ, R. (1976). A perceptual channel for information transfer over kilometer distances: Historical perspective and recent research. *Proc. IEEE*, *64*, 329-354.

Radin, D. I. & Nelson, R. D. (1989). Evidence for consciousness-related anomalies in random physical systems. *Found. Phys. 19(12)*, 1499-1514.

Targ, R. & Puthoff, H.E. (1974). Information transmission under conditions of sensory shielding. *Nature 252*, 602-607.

Tegmark, M. (2000). Importance of quantum decoherence in brain processes. *Phys. Rev. E 61*, 4194-4206.

Wackermann, J., Seiter, C., Keibel, H. & Walach, H. (2003). Correlations between brain electrical activities of two spatially separated human subjects. *Neuroscience Letters 336*, 60-64.

Address for correspondence: Gao Shan,
The Scientists Work Team of Electro-Magnetic Wave Velocity, Chinese Institute of Electronics, LongZeYuan 24-3-501, Hui Long Guan, ChangPing District, Beijing 102208, P.R. China.
E-mail: rg@mail.ie.ac.cn

PARANORMAL BELIEF, RELIGIOSITY AND COGNITIVE COMPLEXITY

Wai-Cheong Carl Tam[1] & Yung-Jong Shiah[2]
[1]*Chung Yuan Christian University, Chung-Li, Taiwan*
[2]*Chien-Kuo Institute of Technology, Taiwan, & University of Edinburgh*

ABSTRACT

Many studies have explored the relationship between paranormal belief and religiosity. Some studies have shown that paranormal belief correlates with religiosity, yet many studies have not supported this association. The first goal of the present investigation therefore was to replicate previous findings. It is also of interest to explore the relationship between paranormal belief and religiosity in Chinese society, since it has not been studied so far. Another interesting correlate with paranormal belief is general cognitive ability. General cognitive ability was found negatively correlated with belief in the paranormal. However, general cognitive ability correlates with education, which might appear to be related to cognitive complexity. The relationship between paranormal belief, religiosity and cognitive complexity was explored in this study. Therefore, this study has two purposes. First, the Chinese version of the Revised Paranormal Belief Scale (RPBS) was constructed. Second, the relationship among paranormal belief, religiosity and cognitive complexity was examined. Cognitive complexity was measured by the repertory grid including an 8X8 grid. Eight personal roles, such as self, grandparent, father, mother, sibling, classmate, friend, and boyfriend or girlfriend, and four constructs, including religious belief, religious activities, virtue vs. evil, and afterlife were preset. Each participant provided the other four constructs during the test. The task was to decide the importance of each construct with regard to the relationship between the individual and each personal role by using a 9-point Likert scale. The task sequence was generated randomly according to the computer program. The score generated by the OMNIGRID called the variability of intensity was a measurement of cognitive complexity of the participant. It was obtained as follows: (a) for each pair of constructs, the correlation (COR) and its variance (VAR = COR^2 X 100%) were calculated, and (b) the variability of intensity was the standard deviation of all the VAR values obtained from each pair of constructs. Forty university students in Taiwan completed the RPBS, the Personal Religiosity Scale and the repertory grid individually. Results indicated that the Chinese version of the RPBS had a satisfactory reliability (Cronbach alpha = .88). The construct validity was confirmed from the correlation matrix between the factors of the RPBS and PRS as well. Paranormal belief and religiosity were two different constructs despite some possible overlap, such as scales of traditional religious belief, spiritualism, believing that nature/environment can affect individuals' well-being and fortune, and afterlife (all these factors had five significant correlations with the other scale). There was a significant negative correlation between religious faithfulness and cognitive complexity. The limitations of this study include small sample size (N = 40) and lack of back-translation procedure of the Chinese version of the RPBS.

INTRODUCTION

Paranormal phenomena are "those which, if genuine, would violate basic limiting principles of science" (Tobacyk, 1988, p. 3), and paranormal belief is belief in paranormal phenomena. The research on paranormal belief can be traced back to the interest on investigating the concept of superstition. For example, Scheibe and Sarbin (1965) tried to conceptualize superstition theoretically, and the variables correlated with superstition/supernatural belief such as conservatism (Boshier, 1973), gender (Bhushan & Bhushan, 1986; Blum, 1976), surgical stress (Shrimali & Broota, 1987), and locus of control (Randall & Desrosiers, 1980; Scheidt, 1973) had also been studied. In addition, the relationship between paranormal belief and the Barnum effect has been investigated (Tobacyk, Milford, Springer & Tobacyk, 1988).

One of the correlates with paranormal belief that attracts much attention is religiosity or religious belief, since religion is an important part of human culture and the presence of God/gods have yet to be proved scientifically. A number of studies have shown that paranormal belief correlated with religiosity (e.g., Buhrmann & Zaugg, 1983; Orenstein, 2002; Thalbourne & Hensley, 2001). However, other studies did not support this association (e.g., Ellis, 1988; Rice, 2003). The fisrt goal of the present investigation therefore was to replicate previous findings. Furthermore, the relationship between paranormal belief and religiosity has not been explored in Chinese society as well.

In the recent decade, although there have been studies in finding the personality or psychopathology variables correlating with paranormal belief (e.g., Dag, 1999; Rattet & Bursik, 2001; Wiseman, Greening & Smith, 2003; Wolfradt, 1997), many studies showed an inconsistent result. For example, empirical studies have shown extraversion to have a high association between paranormal belief and alleged paranormal experience (Honorton, Ferrari & Bem, 1992; Schmeidler, 1982). The sheep–goat effect was found in individuals with higher paranormal belief scores (sheep) to be more extraverted than disbelievers (goats) (Thalbourne, 1981). However, some studies indicated extraversion was not associated with paranormal belief (Rattet & Bursik, 2001; Windholz & Diamant, 1974). Consequently, extraversion seems to be not as obvious a predictor of belief in paranormal phenomena. One possible explanation for the failed replications is the limited reliability of this psychological trait. Therefore, one of our attempts in this study is to find a stable psychological indicator of paranormal belief.

Intelligent or highly educated participants have been shown to have less paranormal belief (e.g., Blum & Blum, 1974; Jahoda, 1970; Killen, Wildman & Wildman, 1974), and this relation did not seem to be accounted for by context effects (Smith, Foster & Stovin, 1998). Moreover, Irwin (1993) contended that no correlation between paranormal belief and intelligence was found from several studies. Musch and Ehrenberg (2002) suggested that general cognitive ability might be a critical underlying variable correlating with paranormal belief and, then, they found general cognitive ability negatively correlated with belief in the paranormal. Nevertheless, intelligence or general cognitive ability correlates with education (Kaufman, 1990), which in turn appears to be related to cognitive complexity. According to the Personal Construct Theory proposed by George Kelly and later elaborated further by his student James Bieri, a cognitive complex person has a personal construct system in which the constructs are clearly differentiated, whereas a cognitive simple person has a personal construct system containing constructs that are poorly differentiated (Potkay & Allen, 1986). The measurement of cognitive complexity indicates the degree of differentiation of the personal constructs construed by the participant, that is, the degree of non-overlapping of these constructs. People who had a higher cognitive complexity related positively to their degree of confidence (Adams-Webber, 2003). We suspected that cognitive complexity might be an important predictor of paranormal belief. The effect of cognitive complexity on paranormal belief has yet to be determined.

Therefore, this study has two purposes. First, the Chinese version of the Revised Paranormal Belief Scale was constructed. Second, the relationship among paranormal belief, religiosity and cognitive complexity was examined. It was hypothesized that cognitive complexity negatively correlated with both paranormal belief and religiosity respectively.

METHODS

Participants

Forty university students (15 males and 25 females) of the Department of Psychology at the Chung Yuan Christian University, Taiwan, were recruited to participate in this study. Their ages ranged from 18.42 to 30.75 years, with a mean of 21.02 years. Religious background of the students comprised 5% Catholic/Christian, 15% in Buddhist, 15% in Taoist (Taiwanese folk religion), 2.5% with other religions (not specified), and the rest without religious beliefs. This sample does not represent the proportions of the religious beliefs in the Taiwan population. Informed consent was obtained from each participant.

Instruments and Procedure

Each participant was administered the following tests in sequence: the repertory grid, the Revised Paranormal Belief Scale and the Personal Religiosity Scale. The instruments are described as follows.

1. The repertory grid: The OMNIGRID version of the repertory grid was used (Sewell, Adams-Webber, Mitterer & Cromwell, 1992; Sewell, Mitterer, Adams-Webber & Cromwell, 1991) and an 8X8 grid was adopted. Eight personal roles, including self, grandparent, father, mother, sibling, classmate, friend, and boyfriend or girlfriend, and four constructs, including religious belief, religious activities, virtue vs. evil, and afterlife were preset. Each participant provided the other four constructs during the test. The task was to decide the importance of each construct with regard to the relationship between the individual and each personal role by using a 9-point Likert scale. The task sequence was generated randomly according to the computer program. Each participant had to make 64 responses in total. The score generated by the OMNIGRID called the variability of intensity was a measurement of cognitive complexity of the participant. It was obtained as follows: (a) for each pair of constructs, the correlation (COR) and its variance (VAR = COR^2 X 100%) were calculated, and (b) the variability of intensity was the standard deviation of all the VAR values obtained from each pair of constructs.
2. The Revised Paranormal Belief Scale (RPBS): Tobacyk and Milford (1983) constructed the Paranormal Scale, which was later revised and known as the RPBS (Tobacyk, 1988). It is the most widely used instrument for measuring paranormal belief (Goulding & Parker, 2001). Although Tobacyk (1988) proposed a seven-factor construct for the RPBS, other researchers have suggested a two-factor construct (e.g., Houran, Irwin & Lange, 2001; Houran & Lange, 2001; Lange, Irwin & Houran, 2000) or a four-factor solution (e.g., Hartman, 1999). In this study, the RPBS was translated to Chinese with slight changes on the wording of items 4 and 18 but preserving the original meaning due to cultural reasons.
3. Personal Religiosity Scale (PRS): It is obvious that religiosity has cultural differences (e.g., Haraldsson & Houtkooper, 1996), and thus a scale constructed by Soong and Li (1988) in Taiwan was used. The PRS consists of 37 items and has eight factors. The name and meaning of the factors are as follows: (a) religious belief--has faithful religious belief, (b) nature/environment--believes that nature/environment, such as *fengshui*, can affect individuals' well-being and fortune, (c) virtue vs. evil--God will reward the good and punish the evil, (d) religious activities--engages in religious activities, (e) value/happiness of life--experiences values and happiness in life, (f) afterlife--believes in afterlife and reincarnation, (g) quality of physical substance/the beliefs in Chinese medicine--believes that different physical substance have different quality which affects health, and (h) birth/pregnancy--believes there is a god in charge of pregnancy. The original PRS utilizes a 4-point Likert scale for each item, but in this study a 7-point Likert scale was used for the sake of sameness as the RPBS.

RESULTS

For the RPBS, the internal consistency (Cronbach alpha) was .88, with the corrected item-total correlation varied from .10 to .76. The results of confirmatory factor analysis showed that seven factors were extracted with 72.8% of the total variance accounted for. The scree plot showed that two factors might be the best solution. The correlation matrix between the factors of the RPBS and PRS was shown in Table 1.

Table 1 Correlation Matrix between the Factors of the Revised Paranormal Belief Scale (RPBS) and Personal Religious Scale (PRS)

	RPBS						
	TRAD	PSI	WITCH	SUPER	SPIRIT	EXTRA	PRE
PRS							
BELIEF	.426**	.369*	.213	.292	.596***	.334*	.164
NAT	.403**	.054	.345*	.601***	.445**	.290	.419**
VIRTUE	.478**	-.117	.169	.299	.274	.223	.334*
ACTIV	.510***	.435**	.356*	.036	.417**	.308	.115
VALUE	.052	.072	-.007	-.014	.036	-.017	.151
AFTER	.850***	.439**	.348*	.237	.611***	.486***	.292
PHYSI	.156	.065	-.040	-.016	.173	.012	.264
BIRTH	-.017	.146	.314*	.539***	.449**	.201	.398*

Note. For the factors of the RPBS, TRAD = traditional religious belief; PSI = psi; WITCH = witchcraft; SUPER = superstition; SPIRIT = spiritualism; EXTRA = extraordinary life forms; PRE = precognition. For the factors of the PRS, BELIEF = religious belief; NAT = nature/environment; VIRTUE = virtue vs. evil; ACTIV = religious activities; VALUE = value/happiness of life; AFTER = afterlife; PHYSI = quality of physical substance; BIRTH = birth/pregnancy.
* $p < .05$. ** $p < .01$. *** $p < .001$.

For the variability of intensity (cognitive complex), the correlation with each of the factors of both the RPBS and PRS was not significant except that with the religious belief factor of the PRS ($r = -.517$, $p < .001$) shown in Table 2 and Table 3.

Table 2 Correlation Matrix between the Variability of Intensity and Paranormal Belief Scale (RPBS)

	RPBS						
	TRAD	PSI	WITCH	SUPER	SPIRIT	EXTRA	PRE
Variability of Intensity	-.223	-.200	.003	-.205	-223	-.191	-.202

Table 3 Correlation Matrix between the Variability of Intensity and Personal Religious Scale (PRS) Revised

	PRS							
	BELIEF	NAT	VIRTUE	ACTIV	VALUE	AFTER	PHYSI	BIRTH
Variability of Intensity	-.517**	-.109	-.200	-.172	-274	-.248	-.191	-.202

** $p < .01$

DISCUSSION

Regarding the construction of the Chinese version of the RPBS, it appears that it had satisfactory reliability (Cronbach alpha = .88). The construct validity could be validated from the correlation matrix between the factors of the RPBS and PRS. The correlations that were significant reflected reasonable relationship between the factors from the two scales respectively in the context of the Chinese culture in Taiwan. For example, the factor traditional religious belief of the RPBS had significant correlations with the following factors of the PRS: religious belief, nature/environment, virtue vs. evil, religious activities, and afterlife. This reflected rather accurately the prevailing traditional Chinese religious thinking. Another example was the factor superstition of the RPBS had significant correlations with the factors nature/environment and birth/pregnancy of the PRS.

Results indicated that paranormal belief and religiosity were two different constructs despite of some possible overlap, such as scales of traditional religious belief, spiritualism, believing that nature/environment can affect individuals' well-being and fortune, and afterlife (all these factors had five significant correlations with the other scale, see Table 1). Two factors of the PRS (virtue vs. evil and quality of physical substance) appeared to have no relationship with the RPBS. The lack of relationship between RPBS and the factor "value/happiness of life" of PRS might imply experiencing one's happiness and paranormal belief were two different constructs. The lack of relationship between RPBS and the factor "quality of physical substance" of PRS might indicate beliefs in Chinese medicine and paranormal belief are two different beliefs. This might reflect the cultural difference between the Western and Eastern worlds as well.

Comparing the items of the traditional religious belief factor and religious belief factor of the RPBS and PRS respectively, the main difference was that the latter factor measured religious faithfulness rather than religious belief. It is interesting that the results of this study indicated the more religiously faithful a person was, the less his/her cognitive complexity would be and vice versa. Since education may help increase one's cognitive complexity, it still needs to be verified empirically whether the latter is a mediator between education and religious belief.

Finally, the limitations of this study include small sample size (N = 40) and lack of back-translation procedure of the Chinese version of the RPBS. More research is also needed to investigate the psychometric properties of both the RPBS and PRS. Moreover, the cultural differences in paranormal belief and religiosity are also important issues to be explored in the future.

REFERENCES

Adams-Webber, JR. (2003). Prototypicality of Self and Differentiating among Others In terms of Personal Constructs. *Journal of Constructivist Psychology, 16* (4), 341-347.

Bhushan, R. & Bhushan, L. I. (1986). Superstition among College Students. *Asian Journal of Psychology and Education, 19(4),* 11-16

Blum, S. H. (1976). Some Aspects of Belief in Prevailing Superstitions. *Psychological Reports, 38*, 579-582.

Blum, S. H. & Blum, L. H. (1974). Do's and Dont's: An Informal Study of some Prevailing Superstitions. *Psychological Reports, 35*, 567-571.

Boshier, R. (1973). An Empirical Investigation of the Relationship between Conservatism and Superstition. *British Journal of Social and Clinical Psychology, 12*, 262-267.

Buhrmann, H. G. & Zaugg, M. (1983). Religion and Superstition in the Sports of Basketball. *Journal of Sport Behavior, 6*, 146-157.

Dag, I. (1999). The Relationships among Paranormal Beliefs, Locus of Control and Psychopathology in a Turkish College Sample. *Personality and Individual Differences, 26*, 723-737.

Ellis, L. (1988). Religiosity and Superstition: Are they Related or Separate Phenomena? *Psychology: A Journal of Human Behavior, 25(2)*, 12-13.

Goulding, A. & Parker, A. (2001). Finding Psi in the Paranormal: Psychometric Measures used in Research on Paranormal Beliefs/Experiences and in Research on Psi-Ability. *European Journal of Parapsychology, 16*, 73-101.

Haraldsson, E. & Houtkooper, J. M. (1996). Traditional Christian Beliefs, Spiritualism, and the Paranormal: An Icelandic-American Comparison. *The International Journal for the Psychology of Religion, 6*, 51-64.

Hartman, S. E. (1999). Another View of the Paranormal Belief Scale. *The Journal of Parapsychology, 63*, 131-141.

Honorton, C., Ferrari, D. & Bem, D. (1992). Extraversion and ESP Performance: Meta-analysis and A New Confirmation. In: Henkel, L.A. and Schmeidler, G.R., Editors, 1992. *Research in Parapsychology*, Scarecrow Press, Metuchen, NJ, 35–38.

Houran, J., Irwin, H. J. & Lange, R. (2001). Clinical Relevance of the Two-Factor Rasch Version of the Revised Paranormal Belief Scale. *Personality and Individual Differences, 31*, 371-382.

Houran, J. & Lange, R. (2001). Support for the Construct Validity of the Two-Factor Conceptualization of Paranormal Belief: A Complement to Thalbourne. *European Journal of Parapsychology, 16*, 53-61.

Irwin, H. J. (1993). Belief in the Paranormal: A Review of the Empirical Literature. *Journal of the American Society for Psychical Research, 87*, 1–39.

Jahoda, G. (1970). Supernatural Beliefs and Changing Cognitive Structures among Ghanaian University Students. *Journal of Cross-Cultural Psychology, 1*, 115-130.

Kaufman, A. S. (1990). *Assessing Adolescent and Adult Intelligence*. Boston, MA: Allyn and Bacon, Inc.

Killen, P., Wildman, R. W. & Wildman, R. W. II (1974). Superstitiousness and Intelligence. *Psychological Reports, 34*, 1158.

Lange, R., Irwin, H. J. & Houran, J. (2000). Top-Down Purification of Tobacyk's Revised Paranormal Belief Scale. *Personality and Individual Differences, 29*, 131-156.

Musch, J. & Ehrenberg, K. (2002). Probability misjudgment, Cognitive Ability, and Belief in the Paranormal. *British Journal of Psychology, 93*, 169-177.

Potkay, C. R. & Allen, B. P. (1986). Personality: Theory, Research, and Applications. Monterey, CA: Brooks/Cole Publishing Company.

Orenstein, A. (2002). Religion and Paranormal Belief. *Journal for the Scientific Study of Religion, 41*, 301-311.

Randall, T. M. & Desrosiers, M. (1980). Measurement of Supernatural Belief. Sex Differences and Locus of Control. *Journal of Personality Assessment, 44*, 493-498.

Rattet, S. L. & Bursik, K. (2001). Investigating the Personality Correlates of Paranormal Belief and Precognitive Experience. *Personality and Individual Differences, 31*, 433-444.

Rice, T. W. (2003). Believe it or not: Religious and other Paranormal Beliefs in the United States. *Journal for the Scientific Study of Religion, 42*, 95-106.

Scheibe, K. E. & Sarbin, T. R. (1965). Towards a Theoretical Conceptualisation of Superstition. *British Journal for the Philosophy of Science, 16*, 143-158.

Scheidt, R. J. (1973). Belief in Supernatural Phenomena and Locus of Control. *Psychological Reports, 32*, 1159-1162.

Schmeidler, G. (1982). A Possible Commonality among Gifted Psychics. *Journal of the American Society for Psychical Research, 76*, 53–58.

Sewell, K. W., Adams-Webber, J., Mitterer, J. O. & Cromwell, R. l. (1992). Computerized Repertory Grids. *International Journal of Personal Construct Psychology, 5*, 1-23.

Sewell, K. W., Mitterer, J. O., Adams-Webber, J. & Cromwell, R. l. (1991). OMNIGRID-PC: A new Development in Computerized Repertory Grids. *International Journal of Personal Construct Psychology, 4*, 175-192.

Shrimali, S. & Broota, K. D. (1987). Effects of Surgical Stress on Belief in God and Superstition: An in situ Investigation. *Journal of Personality and Clinical Studies, 3*, 135-138.

Smith, M. D., Foster, C. L. & Stovin, G. (1998). Intelligence and Paranormal Belief: Examining the Role of Context. *The Journal of Parapsychology, 62,* 65-77.

Soong, W. –L. & Li, Y. –Y. (1988). Personal Religiosity: A new Observation of Religious Belief in Taiwan. *Tsing Hua Journal of Chinese Studies, 18*, 113-139. [in Chinese]

Thalbourne, M. (1981). Extraversion and the Sheep–goat Variable: A Conceptual Replication. *Journal of the American Society for Psychical Research*, *75*, 105–119.

Thalbourne, M. & Hensley, J. H. (2001). Religiosity and Belief in the Paranormal. *Journal of the Society for Psychical Research, 65*, 47.

Tobacyk, J. J. (1988). *A Revised Paranormal Belief Scale*. Unpublished manuscript. Louisiana Tech University, Ruston, LA.

Tobacyk, J. & Milford, G. (1983). Belief in Paranormal Phenomena: Assessment Instrument Development and Implications for Personality Functioning. *Journal of Personality and Social Psychology, 44*, 1029-1037.

Tobacyk, J., Milford, G., Springer, T. & Tobacyk, Z. (1988). Paranormal Beliefs and the Barnum Effect. *Journal of Personality Assessment, 52*, 737-739.

Windholz, G. & Diamant, L. (1974). Some Personality Traits of Believers in Extraordinary Phenomena. *Bulletin of the Psychonomic Society*, *3*, 125–126.

Wiseman, R., Greening, E. & Smith, M. (2003). Belief in the Paranormal and Suggestion in the Séance Room. *British Journal of Psychology, 94,* 285-297.

Wolfradt, U. (1997). Dissociative Experiences, Trait Anxiety and Paranormal Beliefs. *Personality and Individual Differences, 23,* 15-19.

Address for correspondence: Wai-Cheong Carl Tam,
Chung Yuan Christian University, 22 Pu-Jen,Pu-chung
Li, Chung-Li, 32023, Taiwan.
E-mail: cheong@cycu.edu.tw

PRECOGNITIVE AVOIDANCE AND PRECOGNITIVE DÉJÀ VU

Daryl J. Bem
Cornell University, New York

At last year's PA convention, I reported on a phenomenon called "Precognitive Habituation" (PH; Bem, 2003), a phenomenon that emerged from a search for a straightforward laboratory demonstration of psi that could: (a) be observed using participants from the general population; (b) be conducted with no instrumentation beyond a desktop computer; (c) be evaluated by simple statistical tests; and, (d) be replicated by any competent experimenter, including a skeptical one.

On each trial of the PH procedure, the participant is shown a pair of negatively arousing or positively arousing (erotic) photographs on a computer screen and asked to indicate which picture of the pair he or she prefers. The computer then randomly selects one of the two pictures to serve as the "habituation target" and displays it subliminally several times. If the participant prefers the picture subsequently designated as the target, the trial is defined as a "hit." Accordingly, the hit rate expected by chance is 50%.

The PH hypothesis is that the repeated exposures of the target can reach back in time to diminish the arousal it would otherwise produce, thereby rendering negatively arousing targets less negative and positively arousing targets less positive. Because the two pictures in each pair are matched for valence and arousal, participants are predicted to prefer the target-to-be on trials with negatively arousing pictures and the non-target on trials with positively arousing pictures. Preferences on trials with non-arousing ("low-affect") pictures were not expected to differ from chance.

More than 400 men and women participated in 9 variations of the PH experiment, including an independent replication by a skeptical investigator. Collectively the studies provided strong support for the two predicted effects. Across the six basic studies, the hit rate was significantly above 50% on negative trials (52.6%, $t(259) = 3.17$, $p = .0008$) and significantly below 50% on erotic trials (48.0%, $t(149) = -1.88$, $p = .031$). Additional replications are currently in progress at several laboratories around the world.

Currently I am using a similar protocol to explore two new precognitive effects: Precognitive Avoidance (PA) and Precognitive Déjà vu (PDV).

METHOD

The Precognitive Avoidance Procedure

On each trial of the Precognitive Avoidance procedure, the participant is again shown two matched pictures and asked to indicate which picture he or she prefers. In this protocol, however, all the pictures are low-affect pictures. The computer then randomly selects one of the two pictures to serve as the trigger for a full-screen exposure of a highly arousing picture; the other picture of the pair produces only a blank screen. If the participant selects the picture that produces the arousing picture, the trial is designated a "hit. The initial PA hypotheses were that participants will selectively avoid exposing themselves to the negatively arousing pictures and, perhaps, selectively expose themselves to the positively arousing pictures. In other words, the "hit" rate will be significantly less than 50% for negative trials and (possibly) significantly greater than 50% for positive trials.

At this point, it appears that there are systematic individual differences: Those high in anxiety show the predicted effects on the negative trials, but those high in sensation seeking show the reverse effect, significantly exposing themselves to the negatively arousing images. Erotic and positive (nonerotic) pictures are not yet showing any systematic patterns. At the moment, there are too few sessions to be confident of these patterns, but there does appear to be precognitive responding with this protocol.

The Precognitive Déjà vu Procedure

The Precognitive Déjà vu experiment is disguised as a test of subconscious eyewitness identification: "In a police lineup, eyewitness are often asked to identify a person whom they have seen only briefly or even only subconsciously. In this study, you will be shown pictures subliminally (i.e., subconsciously) and asked to identify them when they are paired with pictures you have not seen." On each trial, the participant sees several subliminal flashes and is then shown two matched pictures and asked to select the one which seems more familiar. Unknown to the participants, they are actually exposed to the randomly selected target picture *after* rather than before they make their selection. The PDV hypothesis is that participants will select as more familiar those pictures they are *about* to see: *Precognitive* Déjà vu.

A pilot version of this procedure was administered to 20 participants in a weekend conference at the Institute of Noetic Sciences. They showed a significant PDV effect, especially those who were high on the personality trait of Openness to Experience. Accordingly, the computer now administers a brief version of this scale before the experimental trials begin. Participants in this pilot experiment knew it was a test of psi, that the computer-selected picture would flash *after* their selection. We are currently testing both the disguised and the non-disguised versions of this protocol.

The current version of the computer program also administers a test of left vs right brain functioning. Research on eyewitness identification has demonstrated that those who describe their selection process in global, gestalt terms ("His face just seemed to pop out at me.") are more accurate in their identifications than those who describe their selection process in more analytic terms ("I remembered that he had bushy eyebrows"). This is not surprising because facial recognition is generally considered to be a right-brained function. To explore this further, 20 of the 40 trials use pictures of human faces, 10 use pictures of animals, and 10 use pictures of buildings or landscapes.

I developed this protocol, in part, because several researchers who expressed an interest in replicating the Precognitive Habituation experiment were not comfortable using the negative or the erotic pictures with their subject populations; some who were located at colleges or universities doubted that they could get the protocol past their Internal Review Boards. The PDV protocol avoids this problem by using only low-affect pictures.

References

Bem, D. J. (2003). Precognitive Habituation: Replicable Evidence for a Process of Anomalous Cognition. In: S. Wilson (Chair), *Proceedings of Presented Papers. from the Forty-sixth Annual Convention of the Parapsychological Association, 2003* (pp. 6-20), Vancouver, Canada.

Address for correspondence: Daryl J. Bem,
Psychology Department, Uris Hall, Cornell University,
Ithaca, NY 14853-7601, USA.
E-mail: d.bem@cornell.edu

CREATIVITY, SUBJECTIVE PARANORMAL EXPERIENCES AND ALTERED STATES OF CONSCIOUSNESS

Nicola J. Holt, Deborah L. Delanoy & Chris A. Roe
Centre for the Study of Anomalous Psychological Processes
University College Northampton

INTRODUCTION

Between 1962 and 2003, 27 experimental studies have explored the relationship between creativity and psi. These suggest there is something about artistic populations that is psi-conducive. Of ten free-response studies working with artistic populations, four found that compared to non-artists, artists obtained a significantly higher hit rate (e.g., Moss, 1969). The remaining six studies used the ganzfeld paradigm, each obtaining above chance psi-scoring, with hit rates between 30% and 50% (MCE=25%), and an overall hit rate of 40% (e.g., Morris, Summers & Yin, 2003). These artistic populations obtained higher hit rates than the general populations of ganzfeld studies – for which a recent meta-analysis estimates a hit rate of 30% (Bem, Palmer & Broughton, 2001). However, the interpretation that this psi-success is due to the creativeness of the participants is questionable – it is confounded by other potential characteristics of these populations, such as extraversion, self-confidence, open-belief systems or a willingness to introspect, and the 'experimenter effect'.

Artistic involvement alone is not a measure of creativity. Creativity is commonly defined as a process whereby a novel product emerges, something that is original and also valuable or adaptive (e.g., Boden, 1996). Research using psychometric measures of creativity as correlates of psi-success (e.g. divergent thinking) have had mixed success (there are more significant outcomes with non-artistic populations) and contradictory outcomes (suggesting that 'creativity' may be associated with magnitude of psi) (e.g. Roe, Anowarun & Mckenzie, 2001). Creativity is a poly-faceted, heterogenous construct, with low convergent validity between its various measures (Hocevar, 1981). Interpretation of its relationship to psi has been further hindered by this complexity, particularly when different measures of 'creativity' have been used in isolation, as Palmer (1978) noted.

This study sought to clarify the relationship between creativity and psi by asking if different dimensions of creativity relate differentially to the likelihood of having subjective paranormal experiences (SPEs) and to a proclivity to experience altered states of consciousness (ASCs). This expanded upon the work of Kennedy, Kanthamani and Palmer (1994) who found incidence of SPEs to be significantly correlated with rating artistic creativity as an important purpose in life (r=.20, p=.04).

METHOD

Design

A correlational design was adopted, and a principal components analysis (PCA) conducted to identify underlying dimensions of creativity and determine their relationship to ASCs and SPEs. The order of measures presented to participants was counterbalanced. There were two versions, in reversed orders, with the 'drawing task' in the middle to act as a break from questionnaires.

Apparatus and Materials

The Emotional Creativity Inventory, Averill (1999). A 30-item inventory examining three facets of emotional experience: preparedness; novelty; authenticity and effectiveness.

Creativity Checklist, (Griffin & McDermott, 1998). A 54-item checklist, assessing involvement in: visual arts, performance arts, writing, music and domestic crafts. A scientific activities subscale was added, based on the Creative Behavior Inventory (Hocevar, 1979).

Self-perceived creativity and the importance of creativity. A 2-item measure on a 7 point Likert scale, asking: "How creative would you describe yourself to be?"; and "Is involvement in a creative practice an importance purpose in your life?".

Gough's Creative Personality Scale, CPS (1979). A sub-scale of the Adjective Checklist (Gough & Heilbrun, 1983) composed of 12 adjectives antithetical to, and 18 associated positively with, 'creative personality'.

Remoteness of Associations. Based on the model that making remote word associations is related to creativity (e.g. Martindale, 1989; Mednick, 1969), quick responses are made to ten stimulus words by writing an associated, but unusual, response.

Shapes, a drawing task, where nine simple abstract lines are turned into 'objects', based on the divergent thinking model of creativity developed by Guildford (e.g., 1967) and Torrance (e.g., 2000). Responses are scored by: *flexibility*, the total number of different ideas produced: and *originality*, the unusualness of ideas.

The Creative Cognition Inventory, CCI (Holt, 2002). A 29-item scale developed to assess the use of different cognitive styles in the creative process, with seven sub-scales: heightened internal awareness; intuition; linear, goal-directed cognition; playful, absorbed cognition; the use of analogy; and oneiric cognition.

The Assessment Schedule for Altered States of Consciousness, ASASC (van Quekelberghe, Altstotter-Gleich, & Hertwick, 1991), consisting of 97 items and eleven sub-scales (delineated in the first column of *Table 1*), including SPEs (extrasensory perception and direct mental influence on living systems).

Barron's Ego-strength Scale, (1968), updated MMPI-2 version (Schuldberg, 1992). A 45-item scale measuring personality integration, self-adequacy and personal adaptiveness and effectiveness.

Participants

211 participants (108 females and 101 males), aged between 18 and 70+ were recruited through opportunity sampling (the UCN psychology participant pool, UCN staff, SPR members and posters/flyers at conferences, art galleries and libraries) and word of mouth (through contacts with creative groups and academics at other universities). Participants included 36 professional artists (e.g. fine artists, poets, composers, film makers) and 28 professional scientists (physicists, chemists and engineers).

Procedure

When an enquiry was made, potential participants were given by hand, emailed or posted a questionnaire with an introductory letter and instructions about the study. Hence participants could select their preferred environment to complete the questionnaire. This could be returned either via a freepost address or by email, for which participants were given an eight-week time scale. Participants were asked to spend only five minutes on 'Shapes' and to complete the word association task as quickly as possible.

RESULTS AND DISCUSSION

From the PCA seven oblique components of creativity emerged, as delineated in the top row of *Table 1*. SPEs correlated most highly with the component labelled 'intrapersonal awareness'(r=.45, p=.000001). This component loaded predominantly on emotional creativity and the sub-scales of the CCI that focused on heightened internal awareness (e.g. paying attention to visual imagery and emotions) and non-linear cognition, and was positively correlated with artistic writing and the visual arts. Involvement in 'music and performance arts' and 'crafts and visual art' correlated at low levels with SPEs, but were not significant when corrected for multiple analysis (r=.19, p=.007 and r=.15, p=.03 respectively).

Table 1: Partial correlations between creativity components and ASCs controlling for ego-strength, gender and age*

	Intrapersonal awareness	Artistic creative persona	Figural divergent thinking	Scientific creative personality	Music and performance arts	Crafts and visual art	Writing
Extraordinary mental processes	***.51***	*.18*	.08	.02	*.14*	.09	-.05
	.000001	*.01*	.28	.80	*.05*	.24	.53
Parapsychological experiences	***.45***	.10	.08	.05	*.19*	*.15*	-.02
	.000001	.18	.26	.45	*.007*	*.03*	.80
Esoterics	***.42***	.10	.09	.03	.13	.04	-.02
	.000001	.15	.23	.63	.07	.55	.81
Positive mystical experiences	***.54***	***.37***	*.20*	*.20*	.10	.11	.08
	.000001	***.000001***	*.005*	*.005*	.15	.13	.26
Negative mystical experiences	***.30***	.11	.11	.02	.02	-.08	.08
	.00001	.13	.13	.79	.75	.28	.28
Imagination	***.55***	***.42***	*.20*	.13	*.17*	*.22*	-.03
	.000001	***.000001***	*.005*	.06	*.02*	*.002*	.66
Dreams	***.55***	***.28***	***.25***	.12	*.19*	*.21*	.08
	.000001	***.0001***	***.0001***	.09	*.009*	*.003*	.25
Dissociation	***.54***	.22	*.17*	*.21*	.10	*.18*	-.11
	.000001	.002	*.02*	*.003*	.16	*.01*	.12
Hallucinations	***.52***	***.27***	.11	.08	.12	.09	.10
	.000001	***.0001***	.12	.28	.09	.20	.18
Hypersensitivity	***.43***	*.19*	.08	.03	*.23*	.04	.07
	.000001	*.01*	.24	.63	*.002*	.56	.31
Changed feelings of time and space	***.45***	***.25***	*.14*	*.20*	*.20*	.08	.05
	.000001	***.0001***	*.04*	*.005*	*.005*	.27	.46
Total score on the ASASC	***.67***	***.33***	*.21*	*.15*	*.19*	.13	.05
	.000001	***.00001***	*.003*	*.04*	*.008*	.07	.47

* Correlations in italics are significant at the p<.05 level; those in bold remain significant when corrected for multiple analysis with the Bonferroni method (p<.0006).

Participants who practiced a 'mental discipline' (e.g., meditation, prayer, yoga, martial arts) scored significantly higher than those who did not on the intrapersonal awareness dimension of creativity only (z=-5.23, p=.0000002). Professional artists scored significantly higher on intrapersonal awareness than both scientists (z=-4.03, p=.00006) and other professions (z=-4.35, p=.00001).

This analysis suggests that of these dimensions of creativity, it is not cognitive flexibility (divergent thinking), creative personality or involvement in particular domains that relate to the reporting of SPEs in this sample, but an openness to and exploration of 'psychological space'. This concurs with the idea that people who have 'internal sensitivity' are more likely to have psi experiences (Honorton, 1972). Although artistic populations may be more likely to possess such a 'cognitive style', mere involvement in artistic domains alone does not appear to be related to SPEs to a significant degree. However, other dimensions of creativity correlate at higher levels with ASC proclivity, hence these may mediate psi-outcome in studies that seek to manipulate states of consciousness as part of a 'psi-conducive' protocol. The next stage of this research will be to explore how well these components of creativity predict psi-success in an experimental paradigm.

References

Averill, J. (1999). Individual differences in emotional creativity: Structure and correlates. *Journal of Personality,* 67, 331-371.

Barron, F. (1968). *Creativity and personal freedom*. Princeton & London: D. Van Nostrand Company.

Bem, D., Palmer, J. & Broughton, R. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology*, 65, 207-18.

Boden, M. (Ed.). (1996). *Dimensions of creativity*. London: The MIT Press.

Gough, H. (1979). A creative personality scale for the adjective checklist. *Journal of Personality and Social Psychology,* 37, 1398-1405.

Gough, H., & Heilbrun, A. (1983). *The adjective checklist manual (1983 edition).* Palo Alto, CA: Consulting Psychologists Press.

Griffin, M., & McDermott, D. (1998). Exploring a tripartite relationship between rebelliousness, openness to experience and creativity. *Social Behavior and Personality,* 26, 347-356.

Guildford, J. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450-464.

Honorton, C. (1972). Reported frequency of dream recall and ESP. *Journal of the American Society of Psychical Research*, 66, 369-374.

Kennedy, J., Kanthamani, H. & Palmer, J. (1994). Psychic and spiritual experiences, health and well-being, and meaning in life. *Journal of Parapsychology*, 58, 353-383.

Martindale, C. (1989). Personality, situation, and creativity. In J. Glover, R. Ronning, & C. Reynolds (Eds.), *Handbook of Creativity*, (p. 211-232). New York and London: Plenum Press.

Mednick, S. (1962). The Associative Basis of the Creative Process. *Psychological Review*, 69, 220-232.

Morris, R., Summers, J. & Yim, S. (2003). Evidence of anomalous information transfer with a creative population. *Proceedings of the Parapsychological Association 46th Annual Convention*, 116-131.

Moss, T. (1969). ESP effects in “artists” contrasted with “non-artists”. *Journal of Parapsychology*, 33, 57-69.

Palmer, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed.), *Advances in Parapsychological Research*. London: Plenum Press.

Roe, C., Anowarun, A. & McKenzie, E. (2001). Sender and receiver creativity scores as predictors of performance at a ganzfeld ESP task. *Journal of the Society for Psychical Research*, 65, 107-121.

Schuldberg, D. (1992). Ego-strength revised – A comparison of the MMPI-2 and MMPI-1 versions of the Barron Ego Strength Scale. *Journal of Clinical Psychology*, 48, 500-505.

Torrance, E. (2000). Research review for the Torrance tests of creative thinking figural forms A and B. Bensenville, IL: Scholastic Testing Service, Inc.

Van Quekelberghe, R., Altstotter-Gleich, C. & Hertwick, E. (1991). Assessment schedule for altered states of consciousness: A brief report. *Journal of Parapsychology*, 55, 377-390.

Address for correspondence: Nicola Jane Holt,
Division of Psychology, University College
Northampton, Park Campus, Boughton Green Road,
Northampton NN2 7AL, UK.
E-mail: nicolajaneholt@aol.com

TESTING THE VALIDITY OF TAROT CARDS: CAN WE DISTINGUISH BETWEEN A REAL AND A CONTROL READING?

Itai Ivtsan & Christopher C. French
Anomalistic Psychology Research Unit, University of London

INTRODUCTION

This study juxtaposes two alternative approaches to understanding the results obtained in Tarot reading: the paranormal and the non-paranormal. People consult advisors using different means of divination and claim that the insights they are supplied with are accurate. The assumption behind this experiment was that if this is true, the participant should give higher ratings to a real reading in terms of overall relevance compared with a control reading. In this study, the real reading is one for which the participants choose their own cards compared with the control reading, which consists of randomly chosen cards.

Non-paranormal explanations of the apparent success of divination systems such as Tarot emphasise such factors as the Barnum effect, i.e., the tendency to interpret general statements as applying specifically and accurately to one's own unique circumstances (e.g., Meehl, 1956). Paranormal explanations would suggest that any card drawn in response to an issue is a direct reflection of an inner state. In other words, we unconsciously choose cards in response to the present issue and these cards are a mirror of our deeper understandings (e.g., Angeles, 1987).

Research testing the validity of divination has focused mainly on astrology with very little research regarding the validity of Tarot cards. Blackmore (1983), however, did test the validity of Tarot cards with respect to personality interpretations. Only the first out of three experiments yielded significant results and even then doubts were raised regarding possible flaws in the design.

This experiment also investigated the possibility that believers in the paranormal might be able to accurately distinguish between real and control readings more effectively than non-believers, with believers giving the real reading a higher rating. A non-paranormal explanation might predict that believers would give higher ratings than non-believers to both real and control readings on the assumption that believers may be susceptible to the Barnum effect and that this greater susceptibility was a factor in leading them to become believers in the first place. Although some previous studies have found evidence to support this position (e.g., Glick, Gottesman, & Jolton, 1989), others have failed (e.g., French, Fowler, McCarthy, & Peers, 1998; Tobacyk, Milford, Springer, & Tobacyk, 1988).

METHOD

Participants

Participants were 30 volunteer students (8 male), aged between 18 and 49 years (mean age = 25.3 years, SD = 8.4). Experimenter 1 (E1) was also the Tarot reader, having had ten years experience of using the cards for divination.

Procedure

E1 informed each participant that one reading would be based upon the specific cards chosen whereas the other would be based upon the five cards that happened to end up at the bottom of the deck. Each participant was instructed to shuffle the cards while thinking of a specific issue upon which s/he would like guidance without revealing the issue to E1. The participant next chose five cards and laid them out, face down on the right hand side of the table, in a specific pattern (the *diamond* layout) as instructed by E1. The first card is placed in the centre of the display with the remaining four cards laid out at the vertices of a diamond around it. This constituted the "Real" reading. Each card in this layout is said to relate to the issue in question in a specific way (Angeles, 1987). The backs of the cards were all identical.

Then both the participant and E1 left the room and a third person (Experimenter 2, E2) came into the room and took the five cards from the bottom of the deck and laid them out in the same pattern as the real reading, only on the left. This spread is the control reading. E2 then used a list of random numbers (originally generated by a true random process REG) to determine if the real reading remained on the right side of the table or was swapped with the control reading on the left for each participant. The cards were then positioned in a way that left no clues for E1 or the participant to be able to distinguish between the layouts. This was achieved by ensuring that all the cards were the same way up and that the separation between cards was approximately equal in the two spreads.

Next, E1 and the participant re-entered the room and E1 provided readings based upon both layouts, always starting with the one on the right. After the readings the participant was asked to rate each of the readings by answering eight questions. The questions asked the participant to say to what extent the reading related to the specific issue s/he had in mind. Three questions were about the reading in general (in terms of interest, insightfulness, and helpfulness) and the other five asked specifically about each card. Answers could range from 1 (not at all) to 8 (very much). An overall average across all eight ratings provided the data for the analysis below.

Participants then completed Thalbourne's Australian Sheep-Goat Scale (ASGS). Allocation to Belief Group was done on the basis of a median split of scores on this scale.

Results And Discussion

The possible range of scores on the ASGS was 0 to 36. The mean score for the Non-Believers Group was 6.13 (SD = 4.99) and that for the Believers Group was 19.26 (SD = 3.89). Results were analyzed using a 2 (Reading: Real vs. Control) x 2 (Belief Group: Believers vs. Non-believers) mixed Analysis of Variance. The ANOVA revealed a significant main effect for Belief Group ($F(1, 28) = 6.09$, $p = .02$) with Believers giving higher overall ratings to the readings (6.09) compared to the Non-Believers (5.09). The overall ratings given to Real vs Control Readings did not differ significantly ($F(1, 28) = .15$, not sig.), but the interaction between Belief Group and Reading was highly significant ($F(1, 28) = 7.42$, $p = .011$). As can be seen from Figure 1, the pattern of results obtained did not correspond to what might be predicted upon the basis of either of the general hypotheses outlined. Post hoc *t*-tests revealed that the ratings given to the Real and Control readings did not differ significantly for the non-believers (means of 5.48 and 4.70, respectively; $t(14) = 1.41$, not sig.), but did so for the believers ($t(14) = 2.79$, $p = .014$). However, the believers gave the Control readings *higher* overall ratings (6.61) than the Real readings (5.58). Simple effects analysis further revealed that the groups did not differ with respect to the ratings given to the Real reading but the believers gave a significantly higher rating to the Control rating compared to the non-believers. This intriguing result obviously requires replication and a replication study is currently being planned. The possibility that the result might be an artifact due to an unintentional sampling bias relating to the order in which the Real and Control readings were read is ruled out by the fact that the Real reading was the first one to be read for 16 participants and the Control reading was first for the remaining 14.

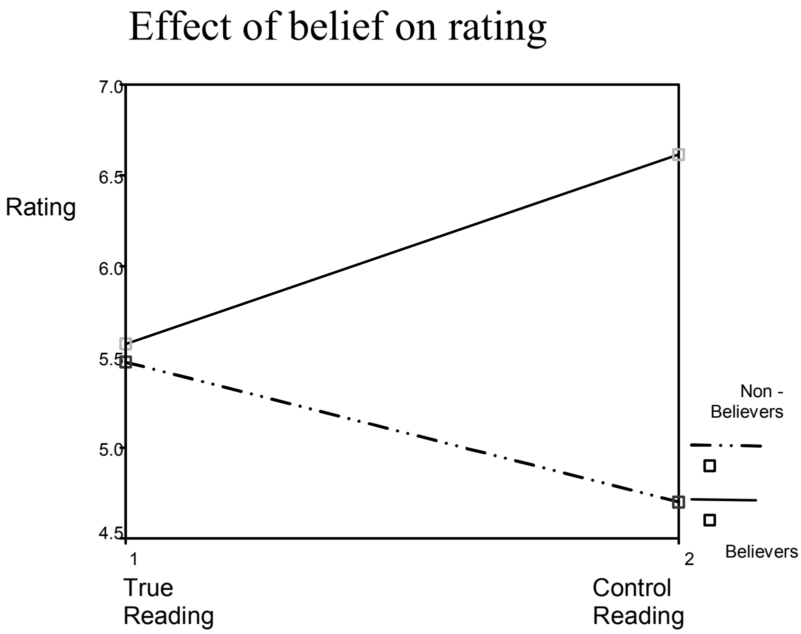


Fig. 1: The effect of Belief Group on the overall rating of the real and control readings.

References

Angeles, A. (1987). *The Tarot Book*. Tel Aviv: Or Am.

Blackmore, S. (1983). Divination with Tarot Cards: An empirical study. *Journal of the Society for Psychical Research. 52,* 97-101.

French, C. C., Fowler, M., McCarthy, K., & Peers, D. (1998). Belief in astrology: A test of the Barnum effect. In K. Frazier (ed.). *Encounters with the Paranormal: Science, Knowledge, and Belief.* Amherst, NY: Prometheus. Pp. 212-220.

Glick, P., Gottesman, D., & Jolton, J. (1989). The fault is not in the stars: Susceptibility of skeptics and believers in astrology to the Barnum effect. *Personality and Social Psychology Bulletin,* 15, 572-583.

Meehl, P.E. (1956).Wanted – A good cookbook. *American Psychologist, 11,* 263-272.

Tobacyk, J., Milford, G., Springer, T., & Tobacyk, Z. (1988). Paranormal beliefs and the Barnum effect. *Journal of Personality Assessment,* 52, 737-739.

Address for correspondence: Christopher C. French, Anomalistic Psychology Research Unit, Goldsmiths College, University of London, New Cross, London SE14 6NW, UK.
E-mail: psa01ccf@gold.ac.uk

THE DREAMS AND VISIONS OF EVA HELLSTRÖM

Stanley Krippner
Saybrook Graduate School, San Francisco

INTRODUCTION

Eva Hellström was born in 1898 and died in 1986, after spending most of her life in Stockholm and helping to form the Swedish Society for Parapsychological Research. Hellström claimed to have had many anomalous experiences over the years, primarily in the form of precognitive dreams and what she referred to as "visions."

Hellstrom recorded these experiences and shared them with serious parapsychological researchers, both in Sweden and abroad. For example, several of these experiences were cited by Ian Stevenson in his 1970 book *Telepathic Impressions* and by Schriever in her 1987 article "A 30-Year Experiment with Time," appearing in *The European Journal of Parapsychology.* Occasionally, Hellstrom described her own experiences, as in "Veridical Precognition of Girls Dancing," published in a 1962 issue of the *Journal of the Society for Psychical Research.* She played an important role in the formation of the Swedish Society for Parapsychological Research, an organization founded in her home in 1947. Until 1968, Hellström served as the Society's secretary and general manager. She initiated the John Bjorkhems Minnesfond Foundation, which she hoped would be devoted to initiating parapsychological research in Swedish universities. However, there were internal disputes within the foundation, primarily about the role to be played by members devoted to spiritualism, a topic that Hellström had abandoned years earlier.

In 1970, Hellström sent me a notebook consisting of 21 of her reported dreams and 45 of her reported visions. They had been neatly typed, and the paper had been glued into the notebook. In an accompanying letter, Hellström wrote me that "these cases are precognitive; most of them have come in deep dreams, some are hypnopompic or hypnagogic visions." Each experience was followed by a lengthy "verification" (pasted into the notebook), typically consisting of statements elicited from other people, from newspaper articles, or from photographs relevant to the purportedly precognized material. Needless to say, these "verifications" were collected after the fact, and conscious or unconscious selection factors could have been at work.

Several dreams and visions were missing from this journal (having been given to other investigators) so I simply worked with the material at hand, asking the question, "What are the characteristics of the reported dreams and visions of Eva Hellström?"

METHODS

To answer this question, I used three instruments, one rating scale, and the services of four judges. Instruments included Krippner and Faith's (2001) 12 "exotic dream" categories (of which 5 were pertinent to this investigation), Casto's Spirituality Scoring System (Casto, Krippner, & Tartz, 1999), Strauch's (2001) content analysis system for "bizarreness," and a 5-unit rating scale to evaluate the purported veridical correspondences between the report and the "precognized" event. Two judged classified the material into the exotic dream categories, one judge evaluated the material on the Casto Scale, a different judge employed the 5-unit rating scale (one=low veridicality; five=very high veridicality), two judges utilized the Strauch scales. In the rare cases when there was a disagreement between judges, I made the final decision.

Casto developed her system in an attempt to identify spiritual content in dream reports. Utilizing the categories from the well-known Hall and Van de Castle system (Hall & Van de Castle, 1966), Casto developed several categories as well as criteria for each, i.e., Spiritual Objects, Spiritual Characters, Spiritual Settings, Spiritual Activities, Spiritual Emotions, Spiritual Experiences.

Strauch developed scales to identify two specific types of bizarre content, "paranormal" and "magical," as well as "miscellaneous bizarre content." Under the heading, "paranormal phenomena," she has differentiated psychological, biological, and physical events. Under the heading of "magical phenomena," she has differentiated magical activities, characters, objects, and scenery. Miscellaneous bizarre content falls into six categories: activities, characters, objects, sensations, settings, and speech. The "paranormal psychological phenomena" category, according to Strauch, comprises extrasensory perception (ESP), psychokinesis (PK), related abilities (e.g., out-of-body sensation, flying), and anthropomorphism (e.g., talking animals). "Paranormal biological phenomena" include bodily changes (e.g., shrinking, becoming taller) and invulnerability. "Paranormal physical phenomena" include animism, changes in form, disappearances, time travel, and unidentified flying objects.

RESULTS

Exotic Dream Categories

Of the 66 dream/vision reports investigated, 55 were classified as "precognitive," 6 as "clairvoyant," 2 as "spirit visitation," 2 as "telepathic," and 1 as "healing." Hellström claimed to specialize in precognitive dreams and the results of this analysis support her claim.

Casto Spirituality Scoring System

Only 4 of Hellström's dream reports were "spiritual" in nature as measured by the Casto Spirituality Scoring System. Dream report #2 contained a spiritual object, "clerical garments and bishop's requisites." Report #79 contained a spiritual character, "The Queen of Heaven." Report #102 contained a spiritual setting, "a cathedral." Report #135 contained a reference to a spiritual character, "a woman in profile with folded hands." In other words, spiritual content was not a predominant theme in Hellström's dreams.

Strauch Scale

Of the 66 dream/vision reports in Hellström's journal, 26 contained "bizarre" elements as defined by the Strauch Scale. Under the subheading "paranormal phenomena," 3 dream reports contained scorable content. Under the heading of "magical phenomena," there was 1 magical object and 4 magical characters. Under the heading, "paranormal phenomena," only three dream reports contained scorable content, one of them a paranormal biological event and two of them paranormal physical events. Under the heading of "magical phenomena," there was one magical object and four magical characters.

"Other bizarre content" includes elements or events that, from an ordinary perspective, are unusual or impossible. This heading included the largest number of entries: 2 bizarre activities, 3 bizarre characters, 5 bizarre objects, and 9 bizarre sensations.

Bizarre sensations were frequently identified in Hellström's dream reports, e.g., auditory and visual hallucinations, feelings of dizziness. Hellström often connected the bizarre sensations with the purported precognized event in her waking life. For example, in case #123, Hellström had a daytime dream in which she hallucinated a bell ringing with no obvious source. When she awakened, she opened her door and discovered a man carrying a large chair up the stairs. He had not rung the bell, as would have been customary.

An example of a "bizarre object" occurred in dream report #59, 6 December 1951, when Hellström and her husband were in Egypt. The bizarre object was water that was "sinking."

> I dreamed last night that I was walking along a big, very wide river; there were floods, and a woman was lying dead in a pool at the riverside. It looked as if the water was sinking, and so she lay there.

When the November 22 Swedish newspaper arrived in Egypt, Hellström read about a disastrous flood in Italy's Po Valley in which "women lost their children and children lost their mothers." This newspaper article was included in Hellstrom's notebook.

A "paranormal biological event" occurred in Case #56, a vision that Hellström recorded on 29 November 1951. It concerned her son, Bosse, who had been studying for an examination.

> I saw a picture of a happy Bosse in a student's cap....When I saw the vision, Bosse looked down at a girl much smaller than himself and he looked so happy and overly delighted, as nobody can look as he can.

In her notes about the vision, Hellström observed that her son wrote his examination papers in the spring of 1952, but failed the verbal section. He scheduled another examination for December 1952, but later postponed it until May 1953. Hellström wrote, "If he manages the verbal part, [he will] make his matriculation and get a student's cap. I believe in success because of my vision." In May, Bosse passed the examination "with distinction" and was overjoyed. He and Eva, the "girl much smaller than himself," were married in 1956. Hellström noted, "This is one of the few happy visions I have had."

Correspondence Rating Scale

Of the 66 reports in Hellström's notebook, 37 received a veridicality rating of one, 7 a score of two, 1 a score of three, 5 a score of four, and 0 a score of five. Apparent veridicality was about the same for both dream and vision reports.

Discussion

A complete collection of Hellström's dream/vision reports would have yielded more complete information than this truncated collection. Nevertheless, Hellström's lengthy accounts of the dreams and visions provide phenomenological insights. Researchers in the field have sorely overlooked this aspect of study. Controlled laboratory studies (e.g., Krippner, Honorton, & Ullman, 1972) are most useful in establishing the veridicality of purported precognitive dreams. However, descriptive studies might discover commonalties of veridical dreams, commonalties that would give experients enough confidence to share a dream that appears to herald an event that has potentially tragic consequences. One can not conclude that these reports were veridical because of the informal nature of the reports and the fact that Hellström herself determined their accuracy. Nevertheless, one must admire her attentiveness to her inner processes, and the care in which she compiled records of these events over the years.

References

Casto, K.L., Krippner, S., & Tartz, R. (1999). The identification of spiritual content in dream reports. *Anthropology of Consciousness, 10*, 43-53.

Krippner, S., & Faith, L. (2001). Exotic dreams: A cross-cultural study. *Dreaming, 11*, 73-82.

Krippner, S., Honorton, C., & Ullman, M.(1972). A second precognitive dream study with Malcolm Bessent. *Journal of the American Society for Psychical Research, 66*, 269-279.

Strauch, I. (2001). Magische und paranomale Phänomene in Träumen [Magical and paranormal phenomena in dreams]. In E. Rüther, A. Gruber-Rüther, & M. Heuser (Eds.), *Träume* [Dreams] (pp. 355-364). Innsbruck, Austria: VIP-Verlag Integrative Psychiatrie.

Address for correspondence: Stanley Krippner,
Saybrook Graduate School and Research Center,
450 Pacific Avenue, 300, San Francisco,
California 94133, USA.
E-mail: skrippner@saybrook.edu

THE MEASUREMENT AND CHARACTERIZATION OF CHARGE ACCUMULATION AND ELECTROMAGNETIC EMISSION FROM BIOENERGY HEALERS

William T. Joines[1], Stephen B. Baumann[1], Jeremy Kim[1],
Jonathan M. Zile[1] & Christine Simmons[2]
[1]*Department of Electrical and Computer Engineering,*
Duke University, Durham
[2]*Rhine Research Center, Durham*

ABSTRACT

In this research project volunteer subjects are asked to focus and direct their mental energy into a region of space. This energy may be in the form of healing intent directed towards another person, nearby or at a distance, or the subject may choose to focus or concentrate their mental energy onto one of several instruments measuring voltage/charge, magnetic fields or electromagnetic emissions across a bandwidth that includes infrared, visible and ultraviolet light. These measurements are conducted in an electrically shielded darkroom with a computer outside the chamber to control data acquisition.

INTRODUCTION

Numerous studies indicate that attentive humans can mentally influence living systems. The range of systems affected runs the gamut from groups of cells (e.g. Nash, 1982); Baumann, et al, 1985) to whole animals (e.g. Grad, 1965; Watkins and Watkins, 1972) to human subjects (e.g. Braud and Schlitz, 1991). Several recent reviews have been published (e.g. Schlitz and Braud, 1997; Targ, 1997; Benor 2001). Benor notes that in 131 controlled studies of distant mental intention, 56 found significant effects. Further, a systematic review of randomized clinical trials found that 57%, involving 2774 patients, showed a positive treatment effect (Astin, et al, 2000).

There are many terms to describe the type of healing this project addresses. Psychic healing, mental healing, laying-on-of-hands, faith healing, energy healing and bioenergy healing are some of the terms used more commonly. The term, bioenergy healing, has become popular in the alternative medicine community. Various techniques are taught for bioenergy healing, such as Reiki, Healing Touch and Therapeutic Touch. Perhaps the most popular among the health care community, and especially among nurses, is Therapeutic Touch (TT). First studied and popularized by Dolores Krieger (1976), a nurse herself, many reviews have taken a cautiously positive interpretation of the bioenergy healing literature (e.g. Achterberg, 1998; Leskowitz, 1998). Benor (2001) reviewed numerous studies showing positive outcomes for patients with a variety of ailments treated with TT.

It is well established in electromagnetic theory that rapid movements of charge can become a source of electromagnetic (EM) radiation (Plonus, 1978). This is the basis of antenna theory and the radiation of radio and television signals. Less well known is the fact that the emission of electromagnetic energy in the form of low-intensity light from biological systems is a common phenomenon. The first reports of biophoton emissions were published eighty years ago by a Russian researcher (Gurwitsch, 1922), and European scientists have studied low-level emission of photons from organisms for quite some time (for reviews see van Wijk, 1992; Tilbury, 1992; Popp, 2002). Of particular importance to our work are reports that human tissues radiate in the ultraviolet region of the spectrum (Rahn, 1936; Konev, et al, 1966). Light emission from humans has been reported in the literature many times over hundreds of years, and an excellent literature review of luminous phenomena around the human body is given by Alvarado (1987).

METHODS

Infrared (IR) emissions are measured using a Probe-Eye Infrared Camera manufactured by Hughes Corporation. The IR camera head is cooled by a small canister of Argon to improve the signal-to-noise ratio. The camera head is located in the inner darkroom for monitoring the subject, and a video display is located outside the darkroom. The display shows the subject outlined in varying shades of color corresponding to temperature. Thus, any movements by the subject that may produce artifacts should be readily observable. These images are also recorded on a VCR.

To record charge build-up and decay that may occur on the skin surface we use a sensitive voltmeter (Keithley Electrometer model 6514) and a switch system (Keithley model 7001) located outside the chamber with an input of 1-18 electrodes from the subject in order to determine the distribution of charge on the body. The number of electrodes used depends upon the comfort level of the subject.

To measure faint amounts of light that may be emitted from the subject we record photons of light using a photomultiplier tube (PMT) in a thermo-electrically cooled housing (Thorn EMI model 9658R). The peak sensitivity occurs near a wavelength of 385 nm and tapers off gradually into the ultra-violet (UV) range. The PMT system includes a power supply and a photon counter.

To measure magnetic fields we will use an F. W. Bell model 7010 gaussmeter with a digital output and resolution of 0.001 Gauss from dc to 20,000 Hz.

A computer is used for recording and displaying in real time the photon counts from the photomultiplier tube, the voltage or charge from the electrodes attached to the subject and the magnetic field intensity from the magnetometer.

RESULTS

In a previous pilot study conducted over a span of several months we repeatedly measured charge build-up and decay, and concurrent light emission from one subject during periods of healing intent. Subsequent to our pilot study, another research laboratory (Green, et al, 1991) measured charge build-up and decay on the body of several subjects during periods of focused intent, but they made no attempt to measure light emission. For the current study we plan to make measurements using at least 12 control subjects and 12 healers. During each experimental session we will also record the output from a random-number generator (RNG) that will be located within the inner darkroom.

DISCUSSION

Bioenergy healers and their patients often report that they feel prickly skin, localized skin temperature increases and energy flowing through their bodies and hands during healing sessions. These subjective sensations may have some physical correlate in unusually high surges in body potential (electrostatic charge build-up and release) that have been measured from some healers. Elmer Green and his associates constructed a room with electrically isolated copper walls and then placed subjects in the room on an insulated platform that isolated them from electrical ground (Green et al., 1991). Surges in body potential measured from 10 experienced meditators did not exceed 4 volts, but in 6 out of 14 experienced, non-contact therapeutic touch practitioners, surges were recorded of 4 to 221 volts (median = 8.3 volts) with durations ranging from 0.5 seconds to 12.5 seconds (median = 3.6 seconds). The frequency of these surges increased during therapeutic sessions in which the practitioner was working on a patient. Video monitoring eliminated the possibility of body motion artifact accounting for the voltage surges. Most of the surges were of negative polarity and returned to baseline within a matter of seconds. Our system has the capability to measure not only voltage or charge fluctuations directly from multiple sites on the bodies of bioenergy healers, but also the ability to measure EM emissions across a wide frequency band that includes IR, visible and UV light.

References

Achterberg J. (1998). Clearing the air in the therapeutic touch controversy. *Alt Ther Health & Med*, *4*, 100-101.

Albrecht-Buehler, G. (1992). Rudimentary form of cellular vision, *Proc. National Academy of Sciences*, *89*(17), 8288-8292.

Alvarado, C.S. (1987). Observations of luminous phenomena around the human body: A Review, *J. Society for Psychical Research*, *54*(806):38-60.

Astin, J.A., Harkness, E., Erns,t E. (2000). The efficacy of "distant healing": a systematic review of randomized trials, *Ann Intern Med*, *132*, 903-910.

Baumann, S., Lagle, J., Roll, W. (1985). Preliminary results from the use of two novel detectors for psychokinesis, *Research in Parapsychology*, 59-62.

Benor, D.J. (2001). *Healing Research Volume 1, Spiritual Healing: Scientific Validation of a Healing Revolution*, Vision Publications, Minnesota.

Braud, W.G., Schlitz, M.J. (1991). Consciousness interactions with remote biological systems: anomalous intentionality effects, *Subtle Energies*, 2(1), 1-46.

Grad, B.R. (1965). Some biological effects of laying-on of hands; a review of experiments with animals and plants. J *Am Soc Psychical Res*, *59*, 95-127.

Green, E.E., Parks, P.A,, Guyer, P.M., Fahrion, S.L., Lolafaye, C. (1991). Anomalous electrostatic phenomena in exceptional subjects, *Subtle Energies*, 2(3), 69-94.

Gurwitsch, A.G. (1922). Uber Ursachen der Zellteilung, *Roux. Archiv*., 52, 167-177.

Gurwitsch, A.G, Eremeyev, V.F., Kababchievsky, Y.A. (1965). Ultra-weak emissions in the visible and ultra-violet regions in oxidation solutions of glycine by hydrogen peroxide, *Nature*, 206, 20-22.

Konev, S,V., Lyskova, T.I., Nisenbaum, G.D. (1966). Very weak bioluminescence of cells in the ultraviolet region of the spectrum and its biological role, *Biofizika*, 11, 410-413.

Krieger, D. (1976). Healing by the 'laying-on' of hands as a facilitator of bioenergetic change: the response of in-vivo human hemoglobin, *Psychoenergetic Systems*, 1, 121-129.

Leskowitz, E. (1998). Un-debunking therapeutic touch. *Alt Ther Health & Med*, 4, 101-102.

Metcalf, W.S., Quickenden, T.I. (1967). Mitogenetic radiation, Nature, 216, 169-170.

Plonus, M.A. (1978). *Applied Electromagnetics*, McGraw-Hill.

Popp, F.-A. (2002). Biophotons – background, experimental results, theoretical approach and applications. *Frontier Perspectives,* 11(1), 16-28.

Rahn, O. (1936). *Invisible Radiations of Organisms*, Verlag von Gebruder Borntraeger, Berlin.

Schlitz, M., Braud, W. (1997). Distant intentionality and healing: assessing the evidence. *Alternative Therapies*, 3(6), 62-73.

Targ, E. (1997). Evaluating distant healing: a research review. *Alternative Therapies*, 3(6), 74-78.

Tilbury, R.N. (1992). The effect of stress factors on the spontaneous photon emission from microorganisms. *Experientia*, 48, 1030-1041.

Van Wijk, R. (2001). Bio-photons and bio-communication, *Journal of Scientific Exploration*, 15(2), 183-197.

Watkins G.K., Watkins, A.M. (1971). Possible PK influence on the resuscitation of anesthetized mice. *JP*, 35, 257-272.

Address for correspondence: Stephen B. Baumann,
Duke University, Department of Electrical and
Computer Engineering, Kingsbury Hall,
Durham, NH 03824, USA.
E-mail: steveb@pstnet.com

COMBINING THE EVENT-RELATED BRAIN POTENTIAL (ERP) AND PRECOGNITIVE HABITUATION TECHNIQUES: A METHODOLOGICAL PROPOSAL

Norman S. Don, Bruce E. McDonough, & Charles A. Warren
Kairos Foundation, Wilmette

The objective of this project was to develop a methodology for combining the event-related brain potential (ERP) technique, used by our group in several past psi experiments, with the newly-developed precognitive habituation (PH) procedure, recently described by Bem (2003). Achieving this would provide a new means for studying the brain processes underlying precognitive psi.

ERPs are minute fluctuations in voltage recorded from the scalp in response to sensory events. Typically observed by averaging the brain's response to a number of similar or repeated stimuli, ERPs are known to reflect various aspects of stimulus processing and cognition. Our past studies indicated that ERPs elicited by target stimuli differ significantly from those elicited by nontarget decoys in forced-choice precognition tasks (Don et al. 1998; McDonough et al. 2002; Warren et al. 1992). We previously suggested that these ERP differences could reflect unconscious detection of the psi targets since the subjects' (conscious) guesses were at chance levels in all of these studies.

The mere exposure, or affective habituation, effect is a well-established phenomenon (Zajonc 1968; Zajonc 2001; Bornstein 1989), whereby the more frequently subjects are exposed to stimuli, the greater their liking for those stimuli. Bem (2003) has presented behavioral evidence, consisting of nine studies with more than 400 subjects, which he interprets as 'precognitive habituation' because a subliminally presented stimulus seems to diminish arousal associated with the same stimulus presented earlier. The fact that the PH effect appears to be stronger when stimuli are presented subliminally, suggests that it works at an unconscious level. Thus, it is reasonable to expect that in the PH task the targets and nontargets might similarly elicit differential ERPs.

The fact that Bem reliably found significant behavioral effects with the PH task, whereas we typically failed to do so, suggests that the PH task might produce even more robust ERP indicators of psi than did our modified Esperciser task (Honorton 1987). Therefore, our objective was to develop a modified version of Bem's PH program which would permit the recording of ERPs to the PH task stimuli.

In the original PH program, each trial starts as two pictorial stimuli, one target and one nontarget, are delivered simultaneously and side-by-side to the left and right of a central fixation point (Bem 2003); the subjects then register their preference for one or the other and continue looking at the screen during several subliminal presentations of the target picture (selected by RNG after the subjects' choice). In order to record unique ERPs to the target and nontarget stimuli, we presently modified the PH program to initially deliver the two pictorial stimuli one-by-one, instead of simultaneously. These are delivered at screen center in random order with a stimulus duration of 150 ms and an interstimulus interval of 2 sec. A synchronizing signal pulse is output from a digital-to-analog device simultaneously with delivery of each of the two pictorial stimuli in order to precisely mark the onset of the stimuli in the EEG record. After this, the remainder of each trial is presented exactly like the original PH program, i.e., the two pictures are presented side-by-side to permit the subjects to register their preference, and the computer-selected target is then presented subliminally several times.

This modification permits examination of brain responses only to the individual presentations of the target and nontarget stimuli delivered at the beginning of each trial; brain activity occurring during other parts of the trial are not considered. Also, in the original PH procedure, a session consists of 48 trials, 12 trials from each of 4 categories of pictures – aversive, erotic, low affect, and fillers, with one target and one nontarget stimulus per trial. Thus, the standard PH task, as modified for ERP recording, would have up to 12 EEG epochs per ERP average, yielding a somewhat low signal-to-noise ratio even before editing for artifact, which may lower it further. Therefore, as a practical matter, it is suggested that future joint use of these two procedures should employ a greater number trials than used in the Bem (2003) study so as to provide at least 20 to 40 artifact-free stimuli in each stimulus category. While this may be accomplished quite simply, it does require that additional stimuli be developed for each category.

The methodology developed here was intended to integrate the ERP and PH procedures, thus enabling future research to address questions such as: In what ways might ERPs to the PH task stimuli resemble, or differ from, those that would obtain in a conventional, habituation study in which ERPs to the test stimuli follow rather than precede the subliminal exposures? Such comparison may provide a means of confirming, or disconfirming, the 'precognitive habituation' interpretation of the observed PH task effects. Another question is whether ERP effects in the PH task might resemble those observed in our previous work, discussed above, i.e., a larger negative slow wave in response to target as compared to nontarget stimuli? Although we did not conceptualize our previous precognition experiment in terms of precognitive habituation, structurally, it was highly similar to the PH task, and so it is quite possible that the same cognitive process was involved. Additionally, while we previously used only low affect stimuli, we may now ask whether ERP effects would also obtain, or perhaps obtain more strongly, for high affect stimuli, i.e., negative or erotic categories? Another obvious question that may be addressed is whether ERP effects in negative and erotic categories will differ from each other? Precognitive habituation seems to have opposite effects on subjects' expressed preferences for negative and erotic stimuli, hence, it is reasonable to hypothesize that opposite, or different, ERP effects also may be observed in response to these two stimulus categories. While these are just a few of the questions that readily come to mind, it is clear that the combined ERP/PH technique developed here provides future researchers with a promising methodology for elucidating the nature of the brain processes underlying precognitive psi.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Dr. Daryl Bem for permitting us to modify the PH program for this purpose and for his kind assistance.

REFERENCES

Bem, D. J. (2003). Precognitive Habituation: Replicable evidence for a process of anomalous cognition. In S. Wilson (Ed.) *The Parapsychological Association 46th Annual Convention: Proceedings of Presented Papers,* Vancouver, 6-20.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin*, *106*, 265-289.

Don, N. S., McDonough, B.E., & Warren, C. A., (1998). Event-related brain potential (ERP) indicators of unconscious psi: A replication using subjects unselected for psi. *Journal of Parapsychology*, *62*, 127-145.

McDonough, B.E., Don, N.S., & Warren, C. A., (2002). Differential Event-Related Potentials to Targets and Decoys in a Guessing Task. *Journal of Scientific Exploration*, *16*, 187-206.

Honorton, C. (1987). Precognition and real-time ESP performance in a computer task with an exceptional subject. *Journal of Parapsychology*, *51*, 291-321.

Warren, C. A., McDonough, B.E., & Don, N. S. (1992). Event-related brain potential changes in a psi task. *Journal of Parapsychology*, *56*, 1-30.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, Monograph Supplement 9 (No. 2), 1-29.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10,* 224-228.

Address for correspondence: Norman S. Don,
Kairos Foundation, 1187 Wilmette Ave, #220,
Wilmette, IL 60091, USA.
E-mail: normdon@uic.edu

RESPONSE CHARACTERISTICS AND LABILITY IN AN EDA-DMILS STUDY USING EMOTIONAL STIMULI

Peter Ramakers
Koestler Parapsychology Unit, Edinburgh

INTRODUCTION

Research in DMILS has been fairly successful over the years (Schlitz and Braud, 1997), but the effect is not well understood. Often the direction of the effect is reversed (Braud, Shafer and Andrews, 1993; Schlitz and Braud, 1997) or the EDA shows a greater differential effect between influence and rest periods, rather than between activate and calm periods, which are often designated as the main experimental conditions. In a re-analysis of data from two DMILS studies (Watt, Ravenscroft and McDermott, 1999; Delanoy, Morris, Brady and Roe, 1999), Stevens (2000) found that the variance of the EDA between influence (calm and activate) and rest periods in the two studies showed a similar, statistically significant, pattern as opposed to the inconsistent results as originally reported for the two studies. Stevens (2000) argues that, because the EDA patterns of those two studies are considerably different and a comparison of these EDA patterns with the results of a study that exposed subjects to weak electromagnetic fields (Stevens, 2001) showed no apparent similarities, the DMILS effect is more likely to be due to a direct influence effect, rather than a transposed response to sensory input. The fact that the EMF study showed a similar pattern when examining variance, i.e. more variance in the EMF periods (equated to influence) than in the control periods (equated to rest), to the two DMILS studies also supports this view.

Stevens (2000) also compared the rest period variances in the two DMILS studies to the variances in the influence periods and found that responders (those with a response greater than 0.2 sigmas at any point after the start of the influence period) had significantly greater mean variances than non responders. Together with findings from Braud and Schlitz (1983) it suggests that people with greater electrodermal lability might be better receivers in a DMILS situation.

The present study was designed to investigate the effects of emotional stimulation of the sender on the EDA of a remote receiver and by also monitoring the EDA of the sender it makes a direct comparison of responses possible. Thus it can be assessed whether the DMILS response is similar to the sender's sensory response, or whether it is indicative of a direct influence (if there is any such influence) as shown by deviations in mean variance across conditions. A comparison between more traditional measures and the variance measure can offer clues as to the possible cause of the responses.

HYPOTHESES:

1) The mean variance in the emotional ("influence") periods will be greater than the mean variance in the neutral ("rest") periods.
2) "Labile" receivers will show a larger difference in mean variance between emotional and control periods when the sender is a "stabile".
3) There will be an interaction between electrodermal type of the sender and receiver.
4) Type of relationship between sender and receiver will have an effect on the difference between mean variances of the emotional and control periods.
5) Intensity of relationship between sender and receiver will have an effect on the difference between mean variances of the emotional and control periods.

METHODS

Participants

90 pairs (10 pairs in a pilot study, 80 pairs in the formal study) of participants will take part. These pairs will consist of males and females (in no particular combination) and their relationship may vary, but the pairs are all recruited by asking one participant to bring a friend or family member.

Apparatus / Materials

The stimuli will consist of a selection of pictures from the International Affective Picture System (IAPS) database. A total of 20 pictures per subject will be used (the same pictures will be shown to all participants, but in a different, randomized order), with 10 pictures around the IAPS mean of 5 for valence and the lowest arousal ratings for the neutral condition, 5 pictures around the highest valence/arousal ratings for the positive condition and 5 pictures with (approximately) the highest arousal rating and the lowest valence rating for the negative condition. Data will be collected according to standard procedures (for the greater part as outlined by Schmidt and Walach, 2000) using two 24 bit serial port model EDA devices, connected to separate, but synchronized pc's. There is also a questionnaire assessing the type (friends, family, etc.) and intensity of relationship.

Procedure

After the purpose of the study is explained to the participants, and the questionnaires filled out, a decision is made as to who will act as sender for the first half of the total session (participants will swap roles between sessions, so that each participant will act as both sender and receiver for one half of the total session). After this, participants will be led to their respective rooms, where the electrodes will be applied and further instructions given. The sender is instructed to immerse themselves (“get into“) each picture to ensure proper mood induction for the duration of the picture presentation (30 s).When both participants are ready the session will start with a relaxation period of 10 minutes in which both the sender and the receiver will hear the sound of waves. A fixation cross is visible on the sender's screen. Then there will be a three minute data collection period prior to picture presentation. The data of these three minutes will be used to determine the electrodermal type of the participant (by calculating the mean variance). After this three minute data collection (at rest) picture presentation will start. Each picture will be presented (in a random order) for 30 s followed by a reaction/recovery period of 15 s, in which the fixation cross will be shown on the screen. Total time per picture will thus be 45 s. Between picture blocks (picture plus fixation cross) there will be a random interval between 0 and 5 s before the next picture will be presented. Each session will last around 25 minutes. After the first session is over, participants will swap roles and receive instructions according to their new roles. Then the second session starts and at the end of that participants are debriefed and the total session ends.

RESULTS

Since this is an ongoing study, no results can be reported as yet, but results will be available for presentation at the convention.

Planned analyses: For hypothesis 1 a t-test between mean *variances* of emotional and control periods will be performed. In addition, a Wilcoxon Signed Ranks Test and PSI scores will be calculated for the *means* of the EDA. Hypotheses 2, 3, 4 and 5 will be analyzed using a 2x2 ANCOVA with type and intensity of relationship as covariates. Post hoc tests will be done where appropriate.

DISCUSSION

The results of this study will hopefully add to our understanding of the DMILS effect, by shedding light on the question whether the DMILS effect is like a transferred sensory response (i.e. a transference of a sender's similar state) or more likely to be as a result of a direct influence. This study will also allow a comparison between different measures of the DMILS effect (those based on mean SCR against analyses of mean variance).

In addition to this, it will add to the database pertaining to the relationship between psi performance and agent-receiver couplings and explore the suitability of emotional states as independent variable in DMILS research.

REFERENCES

Braud, W. G. and Schlitz, M. J. (1983). Psychokinetic influence on electrodermal activity. *Journal of Parapsychology, 47*, 95-119.

Braud, W., Shafer, D. & Andrews, S. (1993). Further studies of autonomic detection of remote staring. J*ournal of Parapsychology, 57*, 391-409.

Delanoy, D. L., Morris, R. L., Brady, C. and Roe, A. (1999). An EDA DMILS study exploring agent-receiver pairing. *Proceedings of the 42nd Annual Convention of the Parapsychological Association*, 68-82.

Schlitz, M. J. & Braud, W. G. (1997). Distant intentionality and healing: Assessing the evidence. *Alternative Therapies, 3*, (6), 62-72.

Schmidt, S. & Walach, H. (2000). Electrodermal activity (EDA) – State-of-the-art measurement and techniques for parapsychological purposes. *Journal of Parapsychology, 64*, 139-163.

Stevens, P. (2000). Human electrodermal response to remote human monitoring: Classification and analysis of response characteristics. *Journal of Parapsychology, 64*, 391-409.

Stevens, P. (2001). Effects of 5 s exposures to a 50μT, 20 Hz magnetic field on skin conductance and ratings of affect and arousal. *Bioelectromagnetics, 22*, 219-223.

Watt, C., Ravenscroft, J. & McDermott, Z. (1999). Exploring the limits of direct mental influence. *Journal of Scientific Exploration, 13*, 515-536.

Address for correspondence: Peter Ramakers, University of Edinburgh, Koestler Parapsychology Unit,
7, George Square, Edinburgh EH8 9JZ, UK.
E-mail: peter@moebius.psy.ed.ac.uk

A QUALITATIVE INVESTIGATION OF SOURCE MONITORING IN THE CONTEXT OF SUBJECTIVE PARANORMAL EXPERIENCES

Christine A. Simmonds
Rhine Research Center, Durham

INTRODUCTION

One of the conclusions of Prospero Lambertini, who was charged with the task of investigating psychic events in the mid 18th Century was that it is difficult for a prophet to distinguish his own thoughts from extrasensory messages. (Jahn, 2002). With that said, the ability to separate memory and imagination from paranormal inputs was described as *essential* in the context of the remote viewing experiments (Puthoff & Targ, 2002). This ability would seem to be an important process for understanding many other paranormal experiences, and of potential relevance for heightening ESP performance in the laboratory. 'Source-monitoring' is the process whereby one discriminates between the sources of a memory or a perception (e.g., Johnson, Hashroudi & Lindsay, 1991). This may be deciding whether imagery is internal in source compared to what is external, deciding on the source of one external stimulus compared with another, or distinguishing between different sources of one's own mental imagery. Source monitoring in the context of parapsychology, would also stretch to a source decision of 'self' versus 'paranormal'. If it is difficult to distinguish source, what sort of cues are employed to make the choice that imagery is from a psi source or from the self?

The usual mechanisms for deriving the source of an impression should be explored before considering what happens in a paranormal context. It is true that people remember and experience the world from two different sources, those derived from the external world, perceptions, and those generated from the internal world of imaginary, reasoning and thought processes (Johnson & Raye, 1981). Both sources produce memories, which in some manner may be considered to be as 'real' as one another; as a record is laid down for both (Johnson & Raye, 1981). There are several cues to the normal process of distinguishing whether an experience derives from a real, perceptual source or an imaginal source summarized in Table 1.

Table 1 Cues employed to distinguish the source of mental imagery (summarized from several authors in Simmonds, 2003)

CUES TO PERCEPTION	CUES TO IMAGINATION
More decision making	Less focused cognitive activity
Reference to concrete reality	Internal subjective space
External objective space	Derived from the self
Defined and full of detail	Diffuse, incomplete and vague
Several sensory modalities	Fewer sensory modalities involved
'Constant' quality	Fleeting and inconsistent
Involuntary	Under voluntary control
Public	Private

Lindsay and Johnson (2000) suggest that source is inferred from a variety of aspects of an experience; perceptual, semantic, and affective content of the thoughts, images and feelings that come into ones mind. They suggest that similar processes are executed for source of memories and perceptions.

Roll (1966) suggested that where ESP is identified from normal memory and imagery processes, it may be because the ESP impression, or impression judged to be from an ESP source lacks the obvious stimulus that ordinary memory traces have associated with them. Therefore, something may be labeled as deriving from a psi source because it is inconsistent with the previous mental or actual activities; it breaks the normal rules. He describes for example how realistic cases of ESP which seem to have a large correspondence with a real event, usually occur in dreams rather than the waking state.

Several authors have noted that at least subjective and possibly objective ESP may occur at the boundary between reality and imagination (e.g., Blackmore & Rose, 1997; Williams, 1997; Simmonds, 2003). In other words, psi is experienced where the usual rules for allocating source are disrupted. 'Psi' is noticed because experiences are unfamiliar, and qualitatively different from the norm. For example, dream imagery that is subjectively realistic may be more likely to be consciously noticed as dreams are usually associated with bizarre symbolic imagery. Likewise, waking thought that is subjectively bizarre may be more noticed as this state is usually associated with logical thinking. Some support for this was demonstrated in a ganzfeld experiment, where there were non-significant trends toward a relationship between subjective mental imagery being judged more 'external' rather than internal in source and both subjective psi and psi performance.

Woofitt's (1992) work on paranormal experiences applied discourse analysis to peoples' narratives of paranormal experiences. This indicated that mundane experiences are often described as preceding the unexpected experience, in a similar manner to the description of traumatic events. As such, a paranormal experience is inconsistent with what just preceded it, and it 'stands out'. It is clear however that both experiencing and explaining tendencies are necessary in a psychic experience. This is supported in Wolfradt's et al. (1999) finding that a combination of analytic and experiential thinking styles, relate to psi experiences, beliefs and perceived abilities.

To date, there have been few systematic and in-depth studies on source monitoring in the context of subjective paranormal experiences[1]. This study explores the phenomena from a qualitative perspective with a range of participants who have differing levels of personal psychic experiences.

METHODS

Design

This research takes a qualitative approach and seeks to explore subjective paranormal experiences with a specific focus on source monitoring of imagery.

Participants

Recruitment was undertaken by word of mouth and advertisement via the Rhine Research Center for people who have had psychic experiences and were willing to be interviewed about them. Participants were mostly recruited from the Rhine Research Center's monthly Paranormal Experiences Group (PEG) and by word of mouth. The PEG is a discussion group of people who have had or who regularly have psychic experiences. Participants with a variety of different forms of experiences were recruited. Between ten and twenty interviews were planned to be carried out for this study.

Data collection

Data collection will take place by means of face-to-face interview at locations and times mutually arranged by the participant and researcher. Interviews will take place in the Rhine Research Center in Durham, North Carolina, or at participants' homes. Interviews will be semi-structured (see appendix for the questions) which allows for consistency across interviews (in terms of set questions) but also for flexibility and for relevant areas of interest and focus to develop according to individual experiences and beliefs. Each interviews will be tape-recorded and transcribed.

[1] Although Boerenkamp (e.g., 1987) for example, has explored paranormal impressions of psychics from a number of perspectives.

PLANNED ANALYSIS

Thematic analysis

The overall approach taken will be phenomenological. As the subject matter concerns subjective paranormal experiences which often encompass spiritual facets, it is considered that a combination of grounded theory (e.g., Glaser, 1992) and transpersonal methods (see Braud & Anderson, 1998) should be utilized to extract themes. Grounded theory allows theory to emerge from the data, rather than taking a hypothesis-testing approach. As such, data will generate greater understanding of the subjective process of distinguishing sources of mental imagery. Interviews will be analyzed by listening intuitively on different levels, paying attention to the emotional reaction of the researcher, the emotional tone of voice of those describing paranormal experiences and to recurring words and themes at the level of content. As a result, themes associated with cues and aspects of experiences for distinguishing what is psi-related to what is not psi related will be generated and written into a report of source monitoring. Transcripts and the analysis will be made available to participants as a check for validity and to allow interviewees to add their perspective into the analysis.

ACKNOWLEDGEMENT

The author would like to thank the PA reviewers for their comments on an earlier draft of this research brief.

REFERENCES

Blackmore. S. J. & Rose. N. (1997). Reality and imagination: A psi conducive confusion? *Journal of Parapsychology, 61*, 321-335.

Boerenkamp, H.G. (1987). A study of paranormal impressions of psychics: VI. Long-term variation in the behaviour of psychics. *European Journal of Parapsychology*, 7, 33-47.

Braud, W. & Anderson, R. (1998).Transpersonal research methods for the social sciences: Honoring *human experience.* Thousand Oaks, California: Sage Publications.

Glaser, B. G. (1992). Basics of grounded theory Analysis: Emergence Vs. Forcing., Mill Valley CA: The Sociology Press.

Jahn, R. (2002). Appendix. The persistent paradox of psychic phenomena: An engineering perspective. In C. T. Tart, H. E. Puthoff & R. Targ. *Mind at Large: IEEE symposia on the nature of extrasensory perception.* Charlottesville VA: Hampton Roads (pp. 210-316).

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88* (1), 67-85.

Lindsay. S and Johnson. M. (2000). False memories and the source monitoring framework: A reply to Reyna and Lloyd (1997). *Learning and Individual Differences*, 12, 145-161.

Puthoff. H. E. & Targ. R. (2002). A perceptual channel for information transfer over kilometer distances: Historical perspective and recent research. In In C. T. Tart, H. E. Puthoff & R. Targ. *Mind at Large: IEEE symposia on the nature of extrasensory perception.* Charlottesville VA: Hampton Roads (pp.11-69).

Roll. W. G. (1966). ESP and memory. *International Journal of Neuropsychiatry*, (September-October), 505-521.

Simmonds, C. A. (2003). *Investigating schizotypy as an anomaly-prone personality.* Unpublished Ph.D thesis. University of Leicester/University College Northampton.

Tart, C. T., Puthoff, H. E., & Targ, R. (2002) (Eds.). *Mind at Large: IEEE symposia on the nature of extrasensory perception.* Charlottesville VA: Hampton Roads.

Williams. C. (1997). The Role of Imagination in the Construction of Anomalous experience. Unpublished PhD Thesis. Edinburgh University.

Wolfradt, U., Oubaid, V., Straube, E. R., Bischoff, N., & Mischo, J. (1999). Thinking styles, schizotypal traits and anomalous experiences. *Personality and Individual Differences, 27*, 821-830.

Wooffitt. R. (1992). Telling tales of the unexpected: The organisation of factual discourse. Harvester Wheatsheaf.

Address for correspondence: Christine A. Simmonds,
Rhine Research Center, 2741 Campus Walk Avenue,
Building 500, Durham, North Carolina, 27705, USA.
E-mail: christine@rhine.org

APPENDIX

Interview questions.

- What does ‘psychic’ mean to you?
- Can you describe the types of experiences that you have?
- For you, what is it like to have an experience?
- What sort of conditions, state of consciousness, etc. are associated with your having an experience? E.g., Can you choose to have an experience?
- Can you describe how your psychic experiences differ from your normal thoughts/feelings/sense of self?
- How is your sense of self maintained during a psychic experience?
- Do you know *in the moment* that you are having a psi-influenced image or is the realization *after* the event?
- Is there anything about the “imagery” that makes it stand out from your normal imagery? If so, how is it different?
- Have you ever had an experience that was unwelcome or that intruded upon you? If so, how did it feel?
- Do you practice any means of protecting yourself from unwanted psi?
- Have you ever felt that a psychic influence became part of you? What was that like? Did it persist after the event?
- How do you realize that something is not coming from you?

EXPERIMENTER EFFECTS AND PSI PERFORMANCE USING A DIGITAL AUTOGANZFELD SYSTEM

Matthew D. Smith & Louie Savva
Parapsychology Research Group, Department of Psychology
Liverpool Hope University College

The ganzfeld procedure has been used within parapsychology for over 30 years as a controlled method of studying 'psi' experiences in a laboratory setting. The utility of the ganzfeld paradigm in providing replicable empirical support for the existence of an anomalous process of information transfer continues to be debated in both the parapsychological literature as well as more 'mainstream' psychological literature (e.g., Bem & Honorton, 1994; Bem, Palmer & Broughton, 2001; Milton, 1999; Milton & Wiseman, 1999; Milton & Wiseman, 2001; Schmeidler & Edge, 1999; Storm, 2000; Storm & Ertel, 2001).

Researchers on both sides of this debate point to the need for further attempts to replicate and extend ganzfeld-ESP findings across a wider range of laboratories and researchers. With this in mind, researchers at Liverpool Hope University College have developed a fully automated computerised ganzfeld testing system that builds upon the strengths of earlier autoganzfeld systems, and which is flexible and easy to install (Fox, Smith & Williams, 2002).

Added to this, the 'experimenter effect' (where some experimenters are consistently more successful than others in obtaining evidence for psi) continues to be a major challenge facing experimental parapsychology (e.g., Kennedy & Taddonio; 1976; Smith, 2003). This is because the experimenter effect is often held responsible for unsuccessful replication attempts. Whilst there is much debate over the replicability of ganzfeld-ESP findings, there has been very little research directly examining the role of experimenter effects in ganzfeld research (e.g., Parker, Miller & Beloff, 1977). This is unfortunate as experimenter variables may play a central role in explaining the replicability of ganzfeld-ESP findings. This is likely to be especially true if experimenter effects are explained, at least in part, in terms of how the experimenter interacts with his or her participants because participants in ganzfeld experiments are tested individually (as opposed to being group tested) and there is sustained experimenter-participant interaction at the beginning of each session and during the important judging stage.

This study will experimentally assess the relative effects of experimenters' attitudes towards psi and expectations of success upon ESP performance by conducting a ganzfeld study with multiple experimenters. Sixteen individuals are being recruited as experimenters on the basis of their prior attitudes towards psi, with the aim of recruiting those obtaining either high or low scores on a measure of attitudes towards psi. Each experimenter will then conduct 8 trials, giving a total of 128 trials. Experimenter expectancy regarding the likely success of the experiment will be manipulated so that half the experimenters are given a positive expectancy of success and half are given a negative expectancy of success. Experimenters in the positive expectancy condition will be told that previous research using the ganzfeld procedure and, more specifically, research using the digital autoganzfeld system has so far been very successful in obtaining strong evidence in support of ESP and that we expected the trials they were about to conduct would obtain similar positive results. Experimenters allocated to the negative expectancy condition will be told that previous research using the ganzfeld procedure has been difficult to replicate. More specifically, they will be told that previous attempts using the digital autoganzfeld system have been unsuccessful in obtaining evidence for ESP and that we expected the trials they were about to conduct would obtain similar negative results.

The effects of these independent variables upon participants' confidence of success and actual performance on a ganzfeld-ESP task will be assessed. This general approach has been used to good effect to examine experimenter expectancy in psychological research (e.g., Rosenthal, 1976) and has been used successfully in some previous research on experimenter effects in parapsychology (e.g., Parker, 1975; Taddonio, 1976; Watt, 2002). However, no previous research has used this approach with the ganzfeld paradigm, nor has any previous research discriminated between a priori attitudes towards psi and more specific expectations about the outcome of the experiment.

The study is using our newly developed digital autoganzfeld system. This system comprises a software application, *DigiGanz*, which is installed onto the hard drives of two Apple iMac computers linked via an Ethernet cable. At Liverpool Hope these two computers are housed in two separate rooms, in different buildings, approximately 30 metres apart. *DigiGanz* allows the two computers to communicate with each other; one acting as the sender's machine, the other as the receiver's machine. The software then co-ordinates the ganzfeld testing procedure by, for example: prompting the experimenter to enter session details; leading both sender and receiver through a relaxation exercise; using a random function to select the target (from a pool of 100 digital movies, each lasting 60 seconds, stored on the computers' hard drives); displaying the target to the sender; guiding the experimenter and receiver through the judging procedure; and storing the data securely. The software also plays white noise to the receiver and records the receiver's mentation. As this system is designed to be relatively easy to use, it is possible to train undergraduate and postgraduate students to use the system fairly quickly. This means that more time and effort can be devoted to training experimenters in other skills deemed to be important in ganzfeld research. These include those social skills associated with creating a warm and friendly atmosphere, such as putting participants at ease, explaining what the experiment involves, and responding to participants' questions. In order to explore more fully the possible effects of the experimenters' attitudes towards psi and expectancy upon the experimenter-participant interaction, video recordings of a sub-sample of ganzfeld testing sessions (two sessions per experimenter) will be made for subsequent content analysis. This will allow us to, for example, examine whether there is a tendency for experimenters in the different conditions to interact with participants in observably different ways. For example, do experimenters given a 'positive' expectancy tend to display more enthusiasm than experimenters given a 'negative' expectancy? Are experimenters' expectations and attitudes towards psi communicated towards participants explicitly or implicitly?

Participants and experimenters will also complete self-report questionnaires that include the Keirsey Temperament Sorter (Keirsey, 1998) (a 70-item self-report personality measure), questions about personal psi experiences, whether they practise any form of mental discipline and their attitudes towards psi phenomena.

PLANNED ANALYSES

Primary analyses

A 2 x 2 between groups ANOVA will be used to assess the effects of experimenter attitudes towards psi ('high' scores versus 'low' scores) and experimenter expectancy ('positive' versus 'negative' expectancy) upon ganzfeld-ESP performance. For this analysis, ratings for targets and decoys for each trial will be converted into standardised ratings. ANOVA will also allow the interaction between these two independent variables to be assessed.

Secondary analyses

These will include a similar 2 x 2 ANOVA examining the effects of experimenters' attitudes towards psi and experimenter expectancy upon participants' confidence of success in the ESP task. Any interaction effect will also be assessed.

Exploratory analyses

We will undertake an exploratory content analysis of the video recordings of a sample (25%) of the ganzfeld testing sessions (the 3rd and 7th session of each experimenter will be recorded). A prototype coding framework will be developed based on previous psychological research on experimenter effects (e.g., Friedman, 1967) in which experimenters' behaviour will be coded in terms of both verbal behaviour (e.g., information-giving; question-asking) and non-verbal behaviour (e.g., eye-contact; smiling). The analysis will explore differences in behaviour between experimenters in the positive and negative expectancy conditions as well as between experimenters with high and low scores on attitudes towards psi.

References

Bem, D. J. & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*, 4-18.

Bem, D. J., Palmer, J., & Broughton, R. S. (2001). Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology*, 65, 207-218.

Fox, J., Smith, M. D., & Williams, C. (2002). Introducing DigiGanz: Describing a digital autoganzfeld system. In Watt, C. (ed.) *Proceedings of the 45th Annual Conference of the Parapsychological Association*.

Friedman, N. (1967). *The Social Nature of Psychological Research*. New York: Basic Books.

Keirsey, D. (1998). Please Understand Me II: Temperament, Character, Intelligence. Del Mar, CA: Prometheus Nemesis.

Kennedy, J. E. & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, 40, 1-33.

Milton, J. (1999). Should ganzfeld research continue to be crucial in the search for a replicable psi effect? Part I. Discussion paper and introduction to an electronic mail discussion. *Journal of Parapsychology*, 63, 309-335.

Milton, J. & Wiseman, R. (1999). Does psi exist? Lack of replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, 125, 387-391.

Milton, J., & Wiseman, R. (2001). Does psi exist? Reply to Storm and Ertel (2001). *Psychological Bulletin*, 127, 434-438.

Parker, A. (1975). A pilot study of the influence of experimenter expectancy on ESP scores. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.) *Research in Parapsychology 1974* (pp. 42-44). Metuchen, NJ: Scarecrow Press.

Parker, A., Millar, B., & Beloff, J. (1977). A three-experimenter ganzfeld: An attempt to use the ganzfeld technique to study the experimenter effect. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 52-54). Metuchen, NJ: Scarecrow Press.

Rosenthal, R. (1976). *Experimenter Effects in Behavioral Research*. Enlarged Edition. NY: Irvington.

Schmeidler, G. R., & Edge, H. (1999). Should ganzfeld research continue to be crucial in the search for a replicable psi effect? Part II. Edited ganzfeld debate. *Journal of Parapsychology*, 63, 335-388.

Smith, M. D. (2003). The role of the experimenter in parapsychological research. *Journal of Consciousness Studies*.

Storm, L. (2000). Research note: Replicable evidence of psi: A revision of Milton's (1999) meta-analysis of the ganzfeld databases. *Journal of Parapsychology*, 64, 411-416.

Storm, L., & Ertel, S. (2001). Does psi exist? Comments on Milton and Wiseman's (1999) meta-analysis of ganzfeld research. *Psychological Bulletin*, 127, 424-433.

Taddonio, J. L. (1976). The relationship of experimenter expectancy to performance on ESP tasks. *Journal of Parapsychology*, 40, 107-114.

Watt, C. (2002). Experimenter effects with a remote facilitation of attention focusing task: A study with multiple believer and disbeliever experimenters. In Watt, C. (ed) *Proceedings of the Parapsychological Association Annual Conference*, Paris.

Address for correspondence: Matthew D. Smith,
Department of Psychology, Liverpool Hope University College, Hope Park, Liverpool, L16 9JD, UK,
E-mail: smithm3@hope.ac.uk

EVENT-RELATED CORRELATIONS BETWEEN BRAIN ELECTRICAL ACTIVITIES OF SEPARATED HUMAN SUBJECTS: PRELIMINARY RESULTS OF A REPLICATION STUDY

Jiří Wackermann, José Raúl Naranjo Muradás, & Peter Pütz
Dept. of Empirical and Analytical Psychophysics,
Institute for Frontier Areas of Psychology and Mental Health (IGPP), Freiburg

Grinberg-Zylberbaum et al. (1994) reported an experiment with pairs of spatially separated, sensorily and electrically isolated subjects, in which one member of the pair was stimulated with light flashes eliciting brain electrical responses (visual evoked potential, VEP) while the other subject was resting; brain electrical activities (EEG) of both subjects were recorded simultaneously. The authors claimed that, if an emotional bond between the subjects was established, 'transferred potentials' (TP) occurred in the electrical activity of the non-stimulated subject synchronously with the VEPs in the stimulated subject. Later attempts to reproduce these findings were only partially successful (Fenwick et al., 1998; Sabel et al., 2001), but positive results from several conceptual replications have been recently reported (Standish et al., 2003; Radin, 2003; Radin, in press). — In our study based essentially on the same paradigm (Wackermann et al., 2003), we have not observed any 'transferred potentials', but we have found significant fluctuations of the average EEG power in the non-stimulated subjects at times of maximal VEPs in the stimulated subjects. The effect was independent from an emotional bond between the subjects. Here we are reporting results of a replication study.

Sixteen pairs of related subjects participated in the study (13M, 19F, age range 19–58 years). The relationship's duration and subjective rating of its intensity were recorded for each participating pair, but no special procedure was used to enhance the emotional connection between them. The subjects were seated in two electrically and acoustically shielded rooms, separated by ~50 cm empty space between the doubled walls of the rooms (average acoustic attenuation −64 dB). One member of the pair (A) was viewing a monitor on which the visual stimuli were intermittently displayed; the other member of the pair (B) was resting, with eyes open, in a dark room. In the stimulation periods of 1 sec duration an alternating black/white checkerboard pattern (three pattern reversals after each 250 msec) was displayed; during the inter-stimulus intervals (ISI) the monitor was blank. Each experimental session consisted of two halves: in one half of the session subject A was visually stimulated, while in the other half the monitor was covered by an opaque shield. The conditions, hereinafter referred to as 'uncovered' and 'covered', were applied in counter-balanced order. 168 stimuli were presented in each condition (duration ~12 minutes each), with ISI's randomly varied from 1.6 to 7.6 sec.

19-channel EEG was recorded simultaneously from both subjects during the sessions (sampling rate 256 data vectors/sec, band pass 0.15–70 Hz), using two EEG recording systems EADC220 (M&I Ltd, Prague, Czech Rep.) and stored on two different computers. Off-line, one-second segments of artefact-free EEG, aligned with the stimulus onset, were visually identified, re-computed against average reference, and de-trended. A pool of N_S post-stimulus and N_I inter-stimulus EEG segments was obtained for each subject and condition. Post-stimulus EEG data were averaged, and effective voltage of the averaged EEG was calculated in a sliding window,

$$V_{\text{eff}}(t) = \sqrt{\frac{1}{l}\sum_{s=-l/2}^{+l/2} u_{t+s}^2},$$

where u_t denotes the averaged EEG voltage at time t (for a given channel and subject), and l is the window length. We used l = 35 data points (i.e. ~136 msec) consistently with Wackermann et al. (2003).

The variable of interest was V^{B}_{eff} of the non-stimulated subject B at latencies t^*, at which V^{A}_{eff} of the subject A in the condition 'uncovered' (i.e., physically stimulated) attained its maximum. To make these values inter-individually comparable, they were divided by V^{B}_{ref} = median of 1000 effective voltages calculated for 1000 subsets of N_S segments which were drawn randomly from N_I inter-stimulus EEG segments and averaged. Thus we evaluated the ratios

$$Q = \frac{V^{B}_{eff}(t^*)}{V^{B}_{ref}}$$

individually for each condition, subjects pair and electrode site. [Note: Re-scaling by V^{B}_{ref} could not affect the results because the statistics relied entirely on intra-individually constructed baselines.]

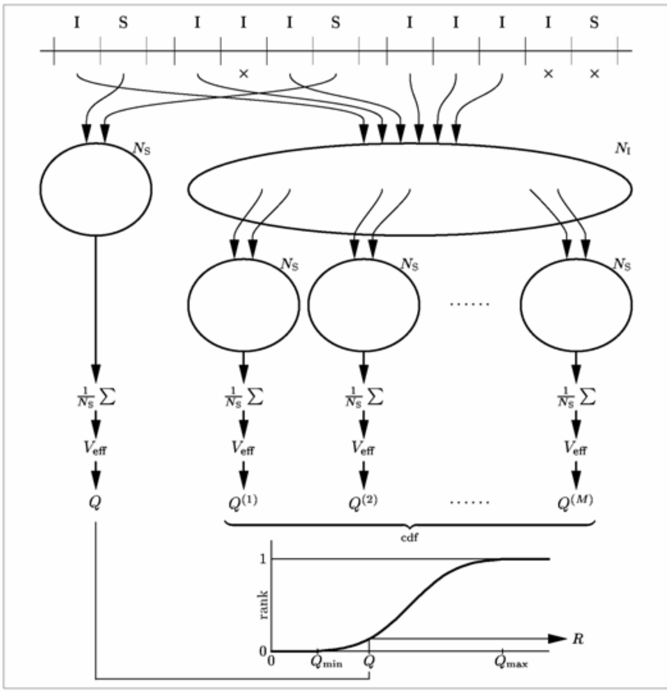


Fig.1. EEG data selection and processing flow chart. S = post-stimulus segment, I = inter-stimulus segment; × indicates data segments containing artefacts and excluded from the analysis. The ovals symbolise the two data pools, consisting of N_S post-stimulus and N_I inter-stimulus EEG segments ($N_S < N_I$).

In each randomisation run, N_S data segments were randomly chosen out of N_I available data segments and processed in the same way as post-stimulus data. The cumulative distribution function of the Q values obtained from M=1000 randomisation runs was constructed and used to transform the post-stimulus Q value to its respective rank R.

According to the null hypothesis there is no difference between the post-stimulus and inter-stimulus EEG in terms of the effective voltage. We used a randomisation statistics to test the hypothesis, as shown in Fig. 1: subsets of N_S segments were drawn randomly from N_I inter-stimulus EEG segments of the subject B, averaged, and their respective Q values calculated; then the rank R of the post-stimulus Q with respect to the cumulative distribution function (cdf) of the inter-stimulus Q's was determined. Given the null hypothesis, the ranks R should be uniformly distributed on the interval [0;1]. To obtain an aggregated score for the whole experimental group, the ranks were transformed to Z-values, averaged across all n=16 subjects, and normalised

$$Z^{\mathrm{rank}} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Phi^{-1}(R_j),$$

where Φ^{-1} denotes the inverse Gaussian cdf; the Z^{rank} values thus should obey the normal Gaussian distribution. These statistics were calculated independently for both conditions 'uncovered' and 'covered' and each electrode location. For a direct comparison between conditions, normalised differences,

$$Z^{\mathrm{diff}} = \frac{Z^{\mathrm{rank}}_{\mathrm{uncovered}} - Z^{\mathrm{rank}}_{\mathrm{covered}}}{\sqrt{2}},$$

were also calculated for each electrode location.

We found predominantly negative Z values in the 'uncovered' and positive Z values in the 'covered' condition. This is a confusing finding: even if we accept a possibility of a communication between the brain states of subjects A and B, it should manifest only in the 'uncovered' (physical stimulation) condition, but not in the 'covered' condition, as there is no electrical correlate of the stimulus in the A's brain. The occurrence of positive Z's in the 'covered' condition may thus indicate an imperfection of our randomisation statistics *or* a true deviation from the baseline distribution. However, if we use the 'covered' condition as the baseline, the distribution of Z^{diff}'s yields a clearer picture: most differences are negative (16 out of 19), five of them significantly deviating from zero ($P < .05$), while none of the positive differences is significant. Fig. 2 shows topographic distributions of Z^{rank}'s for both experimental conditions and Z^{diff}'s for the difference between conditions. Maximal effects are seen in the left parieto-occipital region and in the right frontal region.

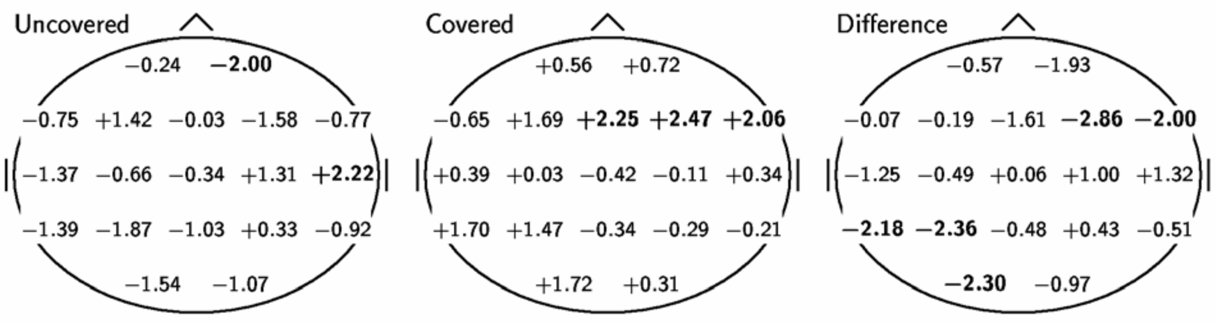


Fig 2: Topographic synopses of Z^{rank}'s and Z^{diff}'s for the nineteen EEG recording locations (system 10/20). Bold typeface indicates Z's significant at the *P*(two-tailed) < .05 level.

Compared to the design of the previous study, the present study employed higher numbers of EEG recording channels, participants and stimuli per session; more broadly varied ISIs; and improved controls, since the 'covered' condition was now applied within each pair instead to another group of participants; [cf. the commentary by Kalitzin & Suffczynski (2003) and the reply by Wackermann (2003); see also Wackermann (2004)]. Our results suggest that, at times of the stimulated subject's maximal response to the visual stimulus, the non-stimulated subject's EEG power was relatively reduced (in terms of the 'uncovered' minus 'covered' difference). Since our earlier findings were inconclusive about the direction and topography of the effect (Wackermann et al., 2003, p. 63), we have undoubtedly made a progress in the characterisation of the effect; however, the interpretation of the effect remains unclear.

Our earlier results suggested that there may be subtle correlations between brain states of two separated subjects (a) caused by a yet unknown mechanism and (b) manifesting themselves when one of the brains responds to an environmental stimulus. However, in the present study we have deviations from mean chance expectation in *both* conditions; such a finding cannot be accounted for by any simple stimulus-response mechanism responsible for biophysical correlations of brain states. We are facing an enigmatic situation, unless we assume that the subject B's brain responds to the physical presence of the stimulus rather than to the subject A's brain response to the stimulus. Even with such an 'ESP-like' interpretation it remains unclear why should the subject B's brain response go to two opposite directions, depending on whether the stimulus has or has not been perceived by subject A. Yet another interpretation might take into account an 'experimenter effect': the experimenters were not exposed directly to the visual stimuli but they were aware of their occurrence. These interpretations imply that our experiment was a kind of unintentional ESP-experiment: an assumption more disturbing than compelling. We hope to obtain more clarity from the next replication study, using a protocol with experimenters being unaware of stimulus presentations, and varying stimulus parameters to modulate the subject A's brain response magnitude.

ACKNOWLEDGEMENT

We wish to thank Samueli Institute for Information Biology for financial support (J.R.N.M., six months), Matthias Gäßler (IGPP) for technical assistance at the experiments, and two anonymous referees for useful comments on an earlier version of this paper.

REFERENCES

Fenwick, P.B.C., Vigus N., & Sanders S. (1998). The transferred potential. London: unpublished manuscript.

Grinberg-Zylberbaum, J., Delaflor M., Attie L., & Goswami A. (1994). The Einstein-Podolsky-Rosen paradox in the brain: The transferred potential, *Physics Essays*, *7*, 422–427.

Kalitzin, S. & Suffczynski, P. (2003). Comments on “Correlations between brain electrical activities of two spatially separated human subjects”. *Neuroscience Letters*, *350*, 193–194.

Radin, D.I. (2003). Thinking outside the box: EEG correlations between isolated human subjects, *Proceedings of the 46th Annual Convention of the Parapsychological Association*, (ed. by S. Wilson), Vancouver, Canada, August 2–4, 2003), pp. 184–199.

Radin, D.I. (2004). Experimental investigation of event-related EEG correlations between isolated human subjects. *Journal of Alternative and Complementary Medicine, 10*, 315–323.

Sabell, A., Clarke C., & Fenwick P. (2001). Inter-subject EEG correlations at a distance – The transferred potential, *Proceedings of the 44th Annual Convention of the Parapsychological Association*, (ed. by C.S. Alvarado), New York, NY, August 2–4, 2001, pp. 419–422.

Standish, L.J., Johnson L.C., Richards T., & Kozak L. (2003). EEG evidence of correlated event related signals between distant human brains, *Quantum Mind 2003*, Tucson AZ: Center for Consciousness Studies, 18–19.

Wackermann, J., Seiter, C., Keibel, H., & Walach, H. (2003). Correlations between brain electrical activities of two spatially separated human subjects. *Neuroscience Letters*, *336*, 60–64.

Wackermann, J. (2003). Correlations between brain electrical activities of two spatially separated human subjects. Reply to the commentary by S. Kalitzin and P. Suffczynski. *Neuroscience Letters*, *350*, 194.

Wackermann, J. (2004). Dyadic correlations between brain functional states: present facts and future perspectives. *Mind and Matter, 2*, 105–122.

Address for correspondence: Jiří Wackermann,
Dept. of Empirical and Analytical Psychophysics,
Institute for Frontier Areas of Psychology and Mental Health (IGPP), Wilhelmstr. 3a, D-79098 Freiburg.
E-mail: jw@igpp.de

MEMORY CONFORMITY AND PARANORMAL BELIEF

Krissy Wilson and Christopher C. French
Anomalistic Psychology Research Unit
Goldsmiths College, University of London

INTRODUCTION

An enormous amount of research has been directed towards understanding general factors affecting the reliability of eyewitness accounts (e.g. Loftus, 1979). Research has shown that memory and perception are active and constructive processes. As a consequence, eyewitness testimony can often be contaminated by incorporation of information presented after the witnessed event. A second strand of research has investigated conformity. Classic research (Asch, 1952) involved asking participants to make a simple perceptual judgement that, under normal circumstances, would always be done correctly. When several stooge respondents gave a clearly erroneous answer before the real participant's turn to respond, many participants gave the wrong answer.

Recent research (e.g., Gabbert, Memon, & Allen, 2003) has drawn these two strands together by showing that if multiple witnesses to the same event discuss what they saw, they may well influence each other's accounts. Gabbert et al. had participants watch a video of a staged crime (a young woman stealing some money) and subsequently tested recall for the event using both free recall and a questionnaire. Participants watched the video either alone or in pairs. Those who watched in pairs were under the impression that each had seen the same video but in fact the videos used were filmed from a slightly different perspective. Some details were therefore visible in the first video (e.g., the woman stealing the money) but not in the second and *vice versa.* Participants were encouraged in the co-witness condition to discuss the event prior to the final recall phase in which it was emphasized that they should only report information that they had actually witnessed with their own eyes. The majority of witnesses (71%) who had discussed the event with a co-witness went on to report items of information that they could not possibly have personally witnessed. Worryingly, 60% of witnesses in the co-witness condition who had not actually witnessed her act of theft were prepared to report that she was guilty of the crime.

The final strand woven into the current research programme is memory recall for ostensibly paranormal events (OPEs). Numerous studies have shown that eyewitness testimony for pseudo-psychic demonstrations, such as fake séances and fork bending, may be inaccurate and vulnerable to memory distortion. Wiseman and Morris (1995), for example, have presented evidence suggesting that believers in the paranormal had poorer memories for "pseudo-psychic" demonstrations (i.e., conjuring tricks) than non-believers. Furthermore, the memory differences between believers and non-believers were particularly marked for information that was crucial to explaining how a particular effect had been achieved. For example, the fact that a key disappeared from view during a metal-bending demonstration was critical because it was at this point that a straight key was switched for a bent key. Believers also tended to rate demonstrations of such pseudo-psychic feats as more "paranormal" than non-believers.

The current research programme aims to bring each of these strands together. It is the first programme to consider the degree to which recall of OPEs can be influenced by the accounts of fellow witnesses, and to also consider whether pre-existing belief in the paranormal is related to susceptibility to such memory conformity effects. The research has obvious and important implications. In both a forensic and anomalistic context, investigators will quite naturally give more evidential weight to accounts from multiple witnesses that broadly appear to mutually support each other. However, it is precisely when multiple witnesses have viewed an unusual real-life incident, such as a crime, or a possible ghost or UFO sighting, that they are likely to discuss the event prior to and during formal investigation. The risk of memory contamination is thus very real and should be taken very seriously by parapsychologists involved in fieldwork or the collection of evidence relating to spontaneous cases.

METHOD

The first study was a direct replication of the Gabbert et al. (2003) paradigm, in an attempt to find memory conformity effects when participants viewed a video of a criminal act. The aims of this study were firstly to investigate the robustness of the effects reported by Gabbert et al. and, secondly, to investigate the possibility that believers in the paranormal might show greater susceptibility to such memory distortion effects even when viewing non-paranormal events. Each member of a dyad watched a video of a staged crime recorded from slightly different viewpoints in such a way that crucial details were not available on both recordings. Each tape included two critical items of information that were not visible on the other tape. Dyads in one condition discussed the events prior to recall while participants in a control condition did not. The first study also assessed whether believers in the paranormal (as assessed by the Belief in the Paranormal Scale; Jones, Russell, & Nickell, 1977) were more prone to such memory distortion even for a non-paranormal event. Participants (N = 40) were assigned to belief group on the basis of a median split.

RESULTS

A chi-square analysis revealed a significant association between experimental condition, i.e., individual witness vs co-witness, and the tendency to report critical (i.e. unseen) items of information in the recall test; (χ^2 (1, 40) = 24.00; $p < 0.001$), with participants in the co-witness condition being significantly more likely to incorporate at least one critical item into their recall compared to controls. Overall, 75% of witnesses in the co-witness condition (i.e., 15 out of 20) reported incorrect items of information that could only have been acquired during the discussion whereas no witnesses in the single witness group did so. Of the 15 participants reporting unseen items of information, 13 reported one item and two reported both items. However, there was no association between memory conformity and belief group (Fisher's exact p = .114). Seven out of ten participants who had not directly witnessed the theft were prepared to say that the young woman was guilty of the crime even though they could personally not have had direct knowledge of this from the videotape.

DISCUSSION

The study found that a significant proportion of witnesses who had discussed an observed event with a co-witness reported items of information that could only have been acquired during the course of that discussion. These results replicate previous findings and demonstrate a robust memory conformity effect. It was not found that believers in the paranormal were more susceptible than non-believers to memory distortion for a non-paranormal event.

A second study is currently running, using the same paradigm but with videotapes of two 2.5-minute clips of a pseudo-psychic performing a selection of apparently psychokinetic demonstrations. Both clips contain essentially the same sequence of events but each includes one important piece of information missing in the other clip, which gives an indication of how the effect was achieved. In the first clip for example, a fork used in a fork bending demonstration is clearly handled by the "psychic" and in the second, the fork clearly goes out of view. The focus of interest is the degree to which the participant's recall is distorted as a direct result of discussions with a co-witness, and whether this relates to pre-existing beliefs in the paranormal.

It is hypothesized that, in line with Wiseman and Morris (1995), in the individual witness condition, believers will be less likely to recall crucial pieces of information compared to non-believers. Since susceptibility to memory distortion correlates with a number of variables that correlate in turn with belief in the paranormal (French, 2003), it is also hypothesized that believers may show a greater tendency for memory conformity in such ostensibly paranormal contexts.

References

Asch, S.E. (1956) Studies of independence and conformity, a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70, 1-70.

French, C.C. (2003) Fantastic memories: The relevance of research into eyewitness testimony and false memories for reports of anomalous experiences. *Journal of Consciousness Studies*, 10, No.6-7, 153-174.

Gabbert, F, Memon, A, & Allan, K (2003) Memory conformity: Can witnesses influence each other's memories for an event? *Applied Cognitive Psychology*, 17, 533-543.

Jones, W. H., Russell, D. W., & Nickell, T. W. (1977). Belief in the Paranormal Scale: An objective instrument to measure belief in magical phenomena and causes. *JSAS Catalog of Selected Documents in Psychology,* 7, 100 (Ms. No. 1577).

Loftus, E.F. (1979) *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Wiseman R. & Morris, R. (1995) Recalling pseudo-psychic demonstrations. *British Journal of Psychology*, 86, 113-125.

Address for correspondence: Christopher C. French,
Anomalistic Psychology Research Unit,
Goldsmiths College, University of London, New Cross,
London SE14 6NW, UK.
E-mail: psa01ccf@gold.ac.uk