European Journal of Parapsychology

Volume 24.1, 2009

European Journal of Parapsychology Volume 24.1 2009

European Journal of Parapsychology

2009
Volume 24.1

European Journal of Parapsychology

Department of Psychology, University of Derby, Kedleston Road, Derby, DE22 1GB
Indexed by PsycINFO.
Printed by JEM Digital Print Services Ltd. on recycled paper.

European Journal of Parapsychology (ISSN: 0168-7263) is published twice a year in Spring and Autumn. Subscription enquiries to subscriptions@ejp.org.uk or postal mail to the Editor at the above address.

INSTRUCTIONS TO AUTHORS

The European Journal of Parapsychology (EJP) is a peer-reviewed scientific journal for research - particularly theoretical and theory-driven empirical work - relating to the field of parapsychology (defined as the study of communication or interaction between organisms and their environment that do not appear to rely on the established sensorimotor channels). The Journal's aim is to stimulate and enhance activity in parapsychology, particularly in Europe, by publishing Articles, Research Notes, Reviews and Comments that offer insight into or criticisms of parapsychological research. The Journal also publishes peer-reviewed Student Research Briefs. Submissions are welcomed on any topic that falls under the heading of parapsychology as defined above. Authors are requested to briefly define any specific terms used (including 'standard' terms such as ESP, PK) in the text of the paper. All manuscripts submitted to the EJP must be under consideration solely by this journal and may not have been previously published in any form (with the exception of translations of non-English articles). Initial submissions should be sent via email to submissions@ejp.org.uk as an attached file (acceptable formats are ASCII, PDF, RTF, OpenOffice or MS Word).

ARTICLES are reports of experimental, clinical or theoretical research, with introductory comments and discussion of results generally being of direct relevance to the data or ideas being presented. Generally articles should be less than 10,000 words in length, though longer papers may be accepted after prior negotiation with the Editors. After evaluation by the editors, the initial article will be sent to two referees for peer-review.

RESEARCH NOTES are manuscripts of up to 3000 words in the following categories: conceptual replications; attempted replications which contradict the original study; reporting on new methodology or apparatus; new applications of existing methodology; challenges to key elements of an existing study, theory or methodology. Research Notes are not intended as a way of publishing inadequate or incomplete work. The Editors or Referees may also decide that a manuscript submitted as an article would be better published as a Research Note.

STUDENT RESEARCH BRIEFS are summaries of research (maximum 2000 words)conducted for assessment as part of a Bachelors or Masters degree (or equivalent) which would not otherwise be submitted as a full article but where the research is of sufficient interest to be disseminated to the wider parapsychological community. Student Research Briefs should provide a summary of methodology and results of the conducted research. They will be subject to peer-review with the understanding that the scope of the research will be less extensive than that of a full article.

REVIEWS provide critical and analytical review or assessment of previously published work. The EJP welcomes novel and cogent reviews (1500 words maximum), subject to Editorial approval.

COMMENTS are extended letters (1000 words maximum) that address specific issues raised in material published in the journal. All comments will be sent for peer-review, the process being similar to article review. Comments will be published in as timely a manner as possible, accompanied where appropriate with a reply from the original author(s).

Full formatting details are provided on the EJP website http://ejp.org.uk/

European Journal of Parapsychology

Volume 24.1

2009

ISSN: 0168-7263

Editorial

Paul Stevens and I took over Editorship of the European Journal of Parapsychology in 2004, our first volume being Volume 19. This marks our tenth issue as the custodians of this journal, and also marks my first as the Editor; Paul Stevens and I having swapped roles.

It also marks the welcome addition of Paul Staples as an Associate Editor for the Journal. I should note that, unlike many journals, we have always treated the different editorial roles as being broadly equal, and we most definitely represent a collective Editorial Team. Most of our Editorial decisions are taken collectively, with much discussion between us and our hard-working Editorial Consultants.

This issue also has our first publication of a Student Research Brief. Paul Stevens and I introduced this classification when we took over the journal, but this is the first paper of its type that we have published. This is a welcome addition, as we have always wanted to encourage undergraduate and Masters students to publish relevant and worthy research. As parapsychology is a small field, students often produce research that is of importance. This classification still adheres to a peer-review process, but the reviewers and readership are informed that this is the work of a student. We hope that undergraduate and early postgraduate supervisors may encourage more of their students to consider publishing worthy and interesting research in this manner.

We would also like to bring to the attention of the readership that from this issue onwards we are revising the publication schedule of the Journal. It will remain two issues per year, but gradually move to Summer (around May) and Winter (around November) issues.

Ian S. Baker, Editor

Paul Staples, Associate Editor

Paul Stevens, Associate Editor

European Journal of Parapsychology
Volume 24.1, pages 5-31

ISSN: 0168-7263

Does Psi Exist?
A Bayesian Approach to Assessing Psi Ganzfeld Data

Michael Y. Lau*, George S. Howard**, Scott E. Maxwell** & Anre Venter**

*Teachers College, Columbia University, New York, USA
**University of Notre Dame, Notre Dame, Indiana, USA

Abstract

Existing meta-analyses in the field of parapsychology on psi Ganzfeld studies have primarily relied on frequentist frameworks of analysis and have yielded conflicting findings. The current study examines the comparative findings from traditional frequentist meta-analysis to Bayesian meta-analysis by adding to the research literature six sets of experiments that contain 120 Ganzfeld trials. Across the six sets of trials, a hit rate of 30% was found when 25% is expected by chance. The addition of the 120 trials to the meta-analytic literature yield findings that are consistent with the originally published meta-analyses. In contrast, the Bayesian meta-analysis yielded findings that depended on strength and magnitude of the priors used. By examining different meta-analytic frameworks, the authors suggest the viability of adopting Bayesian meta-analysis in the parapsychology literature. The relevant issues of the file-drawer problem and a self-corrective science were discussed.

Correspondence details: Dr. Michael Y. Lau, Department of Counseling and Clinical Psychology, Teachers College, Columbia University, 525 W. 120th Street, Box 102, New York, NY 10027. Email: lau@tc.columbia.edu.

Introduction

The empirical study of *psi*, defined as "anomalous processes of information or energy transfer, processes such as telepathy or other forms of extrasensory perception that are currently unexplained in terms of known physical or biological mechanisms" (Bem & Honorton, 1994, p. 4) has primarily existed on the fringes of mainstream psychological research. Recently, beginning with a meta-analysis conducted by Daryl Bem and Charles Honorton (1994) and published in *Psychological Bulletin*, the existence of psi has been thrown in the debate of mainstream psychology (Milton & Wiseman, 1999, 2001; Storm & Ertel, 2001). Central to the debate are the questions, "Does psi exist?" and "To what extent have decades of psi research served to support or refute the existence of it?"

Meta-analyses of Ganzfeld studies

There is a long history of psi research and meta-analytic reviews have been attempted by numerous parapsychologists to examine the aggregated existing literature. Early analyses have mostly suggested the positive existence of psi effects (Honorton, 1985; Hyman, 1985; Rosenthal, 1986). Recent meta-analyses have contributed more conflicting results and heated debates. Bem and Honorton (1994), for example reported a mean effect size[1] (*ES*) of 0.162 (Stouffer $Z = 2.52$, $p = .002$, one tailed) whereas Milton and Wiseman (1999) found a mean *ES* of 0.013 (Stouffer $Z = 0.70$, $p = .24$, one-tailed). Storm and Ertel (2001)

[1] Effect size ($\frac{z}{\sqrt{N}}$, where N is the number of trials in a study), was computed by first obtaining a *z* score for each study included in the meta-analysis. *Z* scores were derived from an exact binomial test when the study measured the outcome by comparing probability of the number of "hits" obtained compared to the number expected by chance. A hit is when a participant correctly identifies the target image sent to him/her by the sender. When more than one outcome was reported in a study, the *z* score associated with the main outcome was used. The mean effect size (*ES*) was obtained by cumulating the effect size scores and dividing that by the number of studies meta-analyzed (Milton & Wiseman, 1999; Storm & Ertel, 2001). The mean $\frac{\Sigma(z/N)}{k}$, where *k* is the number of studies meta-analyzed) is an estimate of *r* (Rosenthal, 1994; Storm & Ertel, 2001).

subsequently published another meta-analysis, with criticisms of Milton and Wiseman's meta-analysis methodology, and found a significant mean *ES* of 0.138 (Stouffer $Z = 5.66$, $p = 7.78 \times 10^{-9}$; see Appendix for more information on these meta-analyses).

Although meta-analysis is promising as a quantitative and integrative methodology for research literatures (Rosenthal & DiMatteo, 2001), inconsistencies between meta-analytic findings render a cumulative scientific enterprise an elusive endeavor (Schmidt, 1992). These inconsistencies in both parapsychology and other areas of psychology are due in part to complex but interrelated factors that include the file drawer problem and the discipline's reliance on a null hypothesis significance testing (NHST) as a dominant statistical analysis framework. All things being equal, studies that produce significant results (in the null hypothesis testing sense; e.g., $p < .05$) are more likely to be published than studies that do not (Rosenthal, 1979), thereby contributing to an overly optimistic estimate of the psi effect in the literature. Also, the "overadoption" (Hubbard, Parsa, & Luthy, 1997) of NHST has received a host of criticisms (Cohen, 1994; Falk & Greenbaum, 1995; Gigerenzer, 1993; Meehl, 1978; Oakes, 1986; Schmidt, 1996) that include opposition to the binary decision making process of either accepting or rejecting the null hypothesis (Cohen, 1994; Folger, 1989; Howard, Fleming & Maxwell, 2000; Rosnow & Rosenthal, 1989) and the misinterpretation of null hypothesis significance testing results (Falk & Greenbaum, 1995; Pollard, 1993). For example, in summarizing the results of an ESP study (Hardy, Harvie & Koestler, 1973), the researchers stated, "Taken together, the receivers scored significantly beyond chance... with a calculated probability of 3,000 to 1 against it being just chance" (p. 117). We argue using empirical data, that adoption of alternative analytical methods not presently adopted in parapsychology research will further researchers' understanding of a controversial topic.

Overview and rationale of present study

Observations and criticisms of statistical practices within parapsychology research (Hyman, 1985; Rosenthal, 1986; Utts, 1986; 1991) mirror those in psychology in general. In light of problematic overreliance on NHST, the goal of the present study is to demonstrate the relative advantages and disadvantages of adopting Bayesian methodology in the analysis of psi Ganzfeld data. Six psi Ganzfeld studies (Delgado-Romero & Howard, 2005; Howard, Lau *et al.*, 2009) with each study consisting of twenty trials. The results were analyzed from NHST, meta-analytic and Bayesian approaches.

Although the 120 trials could have been considered one single study, the focus of this paper is on the methodological approaches of analyzing the data, rather than the final estimate of the psi effect across the trials conducted by this set of researchers. Therefore, an a priori decision was made to conduct enough studies (i.e., six) that were manageable in length and effort (i.e., twenty trials each) to execute the methodological exercise we have chosen to demonstrate below. By creating separate studies from these 120 trials, we hope to illustrate an alternative framework for methodologically considering future psi research evidence.

Method

Ganzfeld procedure

The *Ganzfeld* (meaning "total field") procedure has, for the last 25 years been the dominant research methodology employed in the study of psi (Dalkvist, 2001). The procedure involves sensory deprivation of the participants to enhance the occurrence of psi (Bem & Honorton, 1994; Honorton & Harper, 1974). Two participants are secluded in separate rooms with one participant (the *sender*) instructed to telepathically send a target image to the other participant (the *receiver*). Typically, the procedure uses a pool of static target images grouped in judging sets of 4 images (although more recent studies have also used

dynamic video targets; e.g., Bem & Honorton, 1994). One randomly selected image from each randomly selected set is presented to the sender to "send" to the receiver. At the end of a trial, the judging set is presented to the receiver for evaluation. This set contains the target image and three decoys presented in a random order and instructions are provided to the receiver to correctly select the image that was sent by the sender.

Participants

Participants were students at a mid-sized university located in the Midwest region of the United States of America. The sample consisted of 240 participants (55.8% female). The mean age of the participants is 19.7 years. A total of 120 Ganzfeld trials (consisting of 2 participants each) were conducted over six studies with 20 trials in each study. Participants enrolled in the study by voluntarily signing up their name on a research study announcement board. Participants may or may not know each other. For the reported 120 trials, each pair of participants was only tested once.

Participants completed a questionnaire examining psi-conducive variables when they arrived for the study that gathered information for the following variables, with the first four items taken directly from Honorton (1997).

Belief in psi: Participants responded to the following question, "On a seven-point scale where '1' indicates strong disbelief and '7' indicates strong belief in psi, circle the degree to which you believe in the existence of psi" using a seven point Likert scale. Across all participants, the mean belief in psi rating was 3.49 (SD = 1.43).

Previous personal experiences with psi: Participants responded to the following item, "If you have had experiences which you thought involved psi, which of the following do you feel you have experienced (please circle all that apply)" with the following choices: Telepathy, Clairvoyance, Precognition, and Psychokinesis. Across all participants,

40.8% (*n* = 98) reported no previous personal experiences with psi, 46.2% (*n* = 111) reported one experience, 11.2% (*n* = 27) reported two experiences, and 1.7% (*n* = 4) reported three experiences. None of the participants reported any experiences with psychokinesis.

Previous participation in psi testing: Participants responded with "Yes" or "No" to, "Have you ever participated in formal laboratory testing of psi phenomena?" Only two participants reported affirmatively to this question and neither of them were receivers in their respective trials.

Practice of mental discipline: Participants responded with "Yes" or "No" to, "Have you ever practiced any form of mental discipline, e.g., meditation, biofeedback, hypnosis, relaxation exercises?" Across all participants, 52.5% (*n* = 126) reported having practiced a form of mental discipline.

Art or music background: Participants responded to the following item, "Are you a(n) (please circle all that apply)" with the following choices: Artist and Musician. Across all participants, 88.8% (*n* = 213) identified as an artist, and 75% (*n* = 180) identified as a musician.

Relationship between participants: Participants responded to the following item, "Are you and the other participant" with the following choices: Acquaintances, Close Friends, Related (Family), and none of the above. Experimenters ensured that in each trial, the pair of participants agreed on the descriptor of their relationship. No pairs indicated that they were Related (family), 19.2% (*n* = 23) marked Acquaintances, 31.7% (*n* = 38) marked Close Friends, and 49.2% (*n* = 59) marked none of the above.

Target stimuli

Target stimuli were images obtained from the *National Geographic* website (http://www.nationalgeographic.org/). The images depicted a

variety of subjects including people, nature and animals. Eighty selected images were grouped into 20 judging sets containing four images each. Each set consisted of images that two researchers judged to be dissimilar in content from each other. Both experimenters in the sender and receiver rooms used duplicates of the 20 sets of images. Sets were identified by numbered folders that contained the images, and individual images were each numbered on the back to aid in their selection and presentation to the participants during the trials.

Procedure

Introductory period: The introductory period consisted of the introduction of the participants to each other and to the experimenters to facilitate a warm and welcoming atmosphere. Participants subsequently completed informed consent forms and the self-report measure described earlier.

Experimental conditions set-up: Prior to the beginning of the trial, the pair of participants were assigned at random to the receiver and sender roles by the flip of a coin. One dedicated experimenter was assigned to each of the participants (two experimenters total per trial) and both stayed with the respective participants for the entire duration of the trial. Participants were seated in separate rooms in comfortable recliners and listened to a 10 minute guided relaxation recording prior to the beginning of each trial. The rooms were not sound insulated, and were on opposite sides and ends of a 50 feet long common hallway. Standard Ganzfeld set-up and procedure were adopted, with the receiver being fitted with translucent hemispheric goggles and headphones. A sixty-watt filtered floodlight was placed in front of the receiver and for the duration of each trial, the receiver listened to a recording of white noise.

Randomization and presentation of stimuli. Randomized numbers were generated using a website[2] by an experimenter not

[2] The website was: http://www.randomizer.org/. This websited uses the "Math.random" command of JavaScript programming language within web browsers used to access the website to generate random numbers)

involved in conducting the trials. Identical note-cards identifying randomly selected set numbers (numbers between 1 and 20) were used by the experimenters for each trial. Set numbers on these note-cards were initially hidden by a blank note-card held to it with a rubber band. The two experimenters coordinate visually and verbally in the common hallway to begin timing of the trial and cease any contact with each other until the end of the trial. Next, the experimenter in the sender room revealed the set to be used from the note-cards and then referred to a previously generated random number table (of numbers between 1 and 4) to determine the image within each set that will serve as the target. The experimenter in the receiver room and the receiver did not have factual knowledge of the true target image used until the end of the trial. When presented with the target image, senders were instructed to concentrate on the target image for 30 minutes (*mentation period*) while the receiver was in the Ganzfeld condition.

Mentation period: During the mentation period, the receiver was instructed to verbally report whatever thoughts, images and feelings that occur during the trial into a microphone connected to a two-way radio (Motorola Talkabout T5420) that transmitted the receiver's verbal mentations to the sender in the other room under voice activated mode (i.e., voice is transmitted only when the sender speaks into the microphone). During this time, the experimenter in the receiver room took notes of the oral mentation by the receiver. The sender was asked to focus on conveying the image while listening to the receiver's oral mentation transmitted through the radio. The two-way radio in the sender's room was set to receive mode only (i.e., to transmit, the sender or experimenter would have been required to press a transmit button while speaking into the radio – which did not happen in any of the trials), so that there was only one-way transmission of voice from the receiver to the sender.

Judging period: At the end of the mentation period, the experimenter in the receiver room revealed the judging set that contained the target image using the same note-card procedure by the experimenter in the

sender room 30 minutes earlier. The four images (3 decoys and the target image) were presented in a random order (randomized table of numbers available to the experimenter in the receiver room was created using the same procedure described earlier). The experimenter read the written mentation report back to the receiver and the receiver was instructed to process the mentation experience prior to picking the correct target image. At the end of the judging period, the sender was brought in to the receiver room and all participants were debriefed. At this time, the experimenter in the sender room revealed to the receiver the correct target used during the mentation period.

Results and Analyses

NHST approach and interpretation

Six studies were conducted and the number of correct hits for each study is presented in Table 1. In the first study, 9 of the 20 pairs of participants yielded correct hits. This translates to a 45% hit rate. An exact binomial test is statistically significant, $p = .04$ (one-tailed). All subsequent studies yielded statistically non-significant findings, when considered individually.

It is plausible that after the significant finding in the first study, researchers examining the data in a strict NHST approach might conclude that psi phenomenon exists. Although this is not true for the psi literature, the larger psychological literature is peppered with examples of single experiment studies whose results are used to definitively support or refute the effect of interest. Moreover, it has been suggested that the accumulation of a series of dichotomous accept/reject results do not lend easily to interpretation (Cohen, 1994; Folger, 1989; Howard, *et al*., 2000; Rosnow & Rosenthal, 1989) and is statistically and methodologically problematic (Bushman, 1994; Hedges & Olkin, 1980). With the advent of effect size measures and meta-analytic techniques, focusing on the mere acceptance or rejection of the null hypothesis has become unacceptable by the psychological community. Parapsychology has been especially sensitive to

demonstrating the replicability of psi and has a long history of running multiple studies.

Table 1. Ganzfeld trial hit rates and binomial test results

Study	*N*	Hits	Hit rate	*p* (one-tailed)
1	20	9	45%	.04
2	20	8	40%	.10
3	20	4	20%	.77
4	20	5	25%	.59
5	20	6	30%	.38
6	20	4	20%	.77
1-6	120	36	30%	.12

Meta-analytic approach and interpretation

Figure 1 represents the meta-analytic treatment of the Ganzfeld data presented in Table 1. The figure shows the percentage of correct hits as studies are meta-analytically combined. A 95% confidence interval (CI) was constructed around the hit rate.

The CI after the first study contains the null value of 25%[3]. With the addition of Study 2 to the analysis, we find that the CI no longer contains the null value, suggesting that the 42.5% hit rate after 2 studies is statistically significant. With the addition of the third study,

[3] One might immediately notice that the CI in Study 1 contains the null value of 25%. This is inconsistent with the findings from the NHST section in which Study 1 was found to be statistically significant using an exact binomial test. The contradiction is actually a function of the estimation procedure employed when constructing CIs for dichotomous data. Whereas a binomial test provides an exact probability under the null hypothesis, constructing CIs for binary variables would generally entail an approximation using a standard normal distribution. For this set of data, the *Wald method* for constructing CI was used. Agresti and Coull (1998) have demonstrated that coverage probabilities are poor when using such "exact" methods (especially when sample size is small and/or when the observed proportion deviates from .5) and have recommended alternative methods such as the *adjusted Wald method* or the *score method* in constructing CIs for binomial proportions. Coverage probabilities using the Wald method tends to be overly conservative, resulting in an overly large coverage. This explains why the CI for Study 1 includes the null value in Figure 1, whereas the binomial test from the NHST section yielded a significant finding. The adjusted Wald and score methods were shown to have coverage probabilities close to the nominal confidence level, even when sample size is small and/or when *p* is close to 0 or 1. To be consistent with the analyses in the following section, we have opted for the more common Wald method CIs. When CIs were constructed using one of the recommended methods (Agresti & Coull, 1998), the adjusted Wald method, the CIs for Study 1 and for Study 1 + 2 did not contain the null value and is considered statistically significant. With the inclusion of the each of the remaining four studies, the CIs all contain the null value.

we find that the CI once again contains the null value, bringing the hit rate to a non-significant level. With each addition of the subsequent three studies, the analysis remains non-significant. With the addition of the last study, the hit rate is further reduced to 30%. From the perspective of a meta-analyst, it appears that although we may be approximating the effect size of interest, we are at the same time unsure whether this estimate is significantly a chance occurrence.

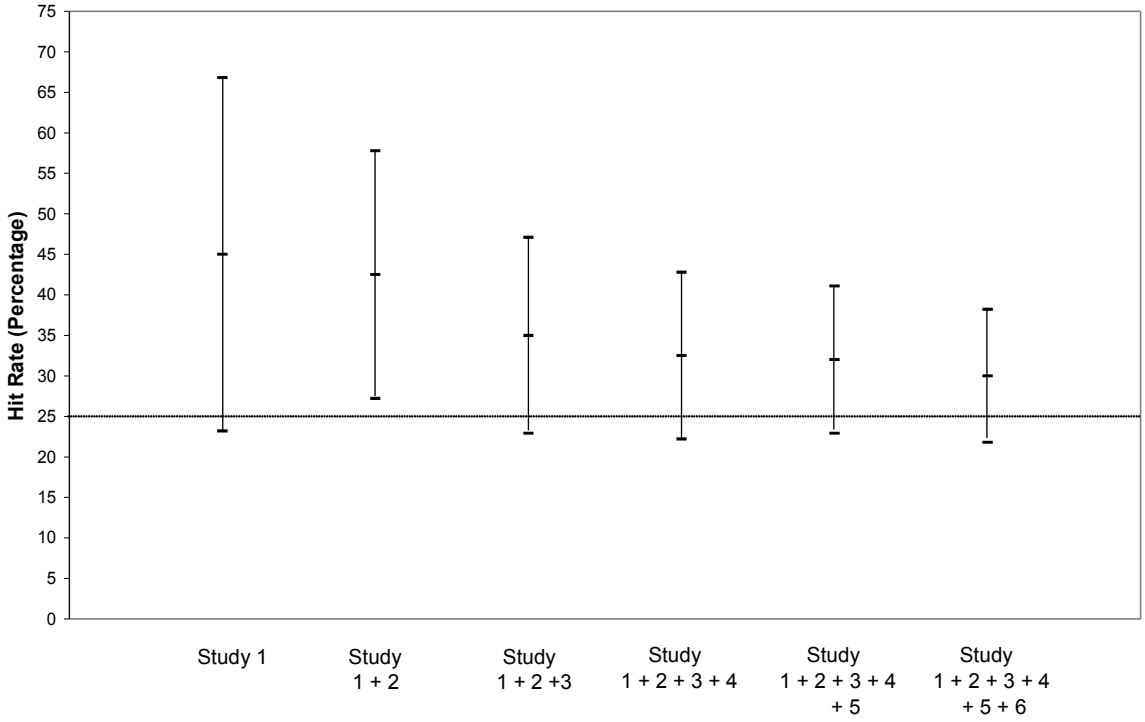

Figure 1. Meta-analysis of the six Ganzfeld studies, with the mean hit rate and 95% confidence interval represented. The dotted line represents the hit rate under the null hypothesis.

The meta-analytic approach is an improvement over the strict NHST approach in several ways. The meta-analytic focuses the analysis on the effect size of interest rather than dichotomous acceptance/rejection of the null hypothesis. Whereas the interpretation of the results from the previous section might conclude that psi phenomena is unlikely to exist given the dominance of non-significant results, a meta-analysis yields an estimate of that effect. When a CI is computed, the meta-analytic approach also yields the statistical

significance at each step of the analysis. It is helpful to note that if the studies were independently graphed, their CIs would be equivalently wide (equal to the length of the CI of Study 1 alone in Figure 1) across all studies. Therefore, the meta-analytic treatment of the data results in greater confidence in effect size estimates with increasing trials.

Bayesian approach and interpretation

Bayesian priors: The Bayesian approach necessitates the specification of an a priori belief for the effect of interest. These so-called *priors* are a probabilistic estimation of personal belief. It may be arrived at by empirical means (such as a meta-analysis of existing literature) or by personal declarations of a belief (which in the absence of a directional belief, is called a "non-informative" prior). To demonstrate how discrepant priors are driven by the data, three priors were constructed using a similar methodology employed by Howard, Maxwell and Fleming (2000). Sixteen graduate student members of the psychology department at the university affiliated with the researchers were surveyed for their personal beliefs in the existence of psi in a Ganzfeld experimental context. The volunteers were provided with the information that in a psi Ganzfeld study a 25% hit rate represented chance occurrence whereas hit rates increasingly higher than the null value represented evidence for psi phenomena. Each student provided: (a) a hit rate estimate of the psi phenomena representing their personal belief, and (b) the number of studies with results in the opposite direction that it would take to change their mind (an estimate of the confidence in (a)). Participants were given the information that each study consisted of 20 trials and that 9 or more trials resulting in hits is statistically significant. The information provided by those surveyed was used to construct a beta distribution defined by [a, b] (Pruzek, 1997). A beta distribution's mean is defined by:

$$\frac{a}{(a+b)} \qquad [1]$$

and its variance by:

$$\frac{ab}{(a+b)^2(a+b+1)} \qquad [2]$$

For example, one of the students surveyed may give a response of 25% hit rate (meaning she does not believe in the existence of psi) and stating that she needs to observe 10 significant studies for her to change her mind about the existence of psi. This translates to a beta distribution of [50, 150][4]. Someone who believes strongly in psi might respond with a hit rate estimate of 50% with 50 null studies for him to change his mind. This is equivalent to a beta distribution of [500, 500]. With more studies needed to change a student's mind, the variance decreases, and the confidence in the estimate increases. The surveyed estimates were subsequently rank ordered, with the median of the top and bottom quartiles representing the priors of "believers" and the "non-believers" of psi. These hypothetical priors are meant to represent extreme prior beliefs for the existence of psi. Finally, a third non-informative prior was included in the analysis as well.

The three priors (believer, non-believer, and non-informative) were used in a Bayesian analysis of the six studies. In Bayesian analyses, priors are combined with results from the data (the *likelihood*) to yield a *posterior probability* (Howard, *et al*., 2000; Pruzek, 1997). In the case of the following analysis, the posterior probability from combining the initial priors with the results from Study 1 becomes the prior probability to be combined with the results from Study 2, and so on. As mentioned earlier, the priors are defined by a beta distribution [a, b]. The likelihood and posterior probabilities are similarly defined by the same parameters. To distinguish among them, let the prior probability be defined by [a', b'], the likelihood probability by [a^*, b^*], and the posterior probability by [a'', b'']. The posterior beta parameters

[4] As indicated earlier, the participants who provided estimates for the priors were instructed to respond in units of studies with each study containing 20 trials. The formula provided by Pruzek (1997) conceptualizes the beta distribution in units of trials. Therefore, the responses made by each participant were multiplied by 20 to yield the parameters to calculate the beta distribution. In this example, the response of 10 significant studies translates to 200 trials, in which 25% are hits, yielding the prior beta distribution of [50, 150].

are simply an additive function of the prior and likelihood parameters (Pruzek, 1997), so that:

$$a'' = a' + a^* \quad [3]$$

and:

$$b'' = b' + b^* \quad [4]$$

Take the non-believer prior as an example. Its beta distribution is defined by [110, 330]. The likelihood distribution for Study 1 is defined by [9, 11]. The posterior beta parameters are therefore defined by [(110 + 9), (330 + 11)], which yields the posterior distribution [119, 341]. The mean of this distribution is 25.8 (Equation 1) and the *SD* is 0.0004 (Equation 2). Table 2 represents the initial prior and the posterior mean hit rates and standard deviation after the inclusion of each of the six studies.

Bayesian approach interpretation: As shown in Table 2, we can see that the initial priors for the non-believers and believers are quite discrepant. The non-believers hold that there are no psi phenomena and that in Ganzfeld studies participants will correctly identify the target image only 25% of the time (at chance level). The hypothetical believers, on the other hand, hold that the hit rate would be at 44.8%, suggesting the existence of psi. In the non-informative prior column, we see that since there is no a priori belief, it does not factor into the posterior probability after the first study. Hence, the values in that column are equivalent to a meta-analytic treatment of the data.

The graphical representation of the Bayesian analysis is displayed in Figure 2. Ninety-five percent confidence intervals were constructed around the mean hit rates (Pruzek, 1997), as defined by

$$\frac{a}{(a+b)} \pm z_{\alpha}/2\sqrt{\frac{ab}{(a+b)^2(a+b+1)}} \quad [5]$$

One notices that not only are the believer and non-believer priors discrepant, their CIs also do not overlap with each other. This is a function of the high confidence in the two extreme views which resulted in relatively tight confidence intervals. Low confidence priors (wide CI) will therefore be more influenced by the data than would higher confidence priors (narrow CI).

Table 2. Mean hit rates and standard deviations for initial prior beliefs and posterior probabilities

	Non-informative Prior		Non-believer Prior		Believer Prior	
	Hit rate	SD	Hit rate	SD	Hit rate	SD
Initial prior	–	∞	25.0%	<1	44.8%	<1
After Study 1	45.0%	11	25.9%	<1	44.8%	<1
After Study 2	42.5%	8	26.5%	<1	44.7%	<1
After Study 3	35.0%	6	26.2%	<1	43.8%	<1
After Study 4	32.5%	5	26.2%	<1	43.2%	<1
After Study 5	32.0%	5	26.3%	<1	42.8%	<1
After Study 6	30.0%	4	26.1%	<1	42.1%	<1

The non-believer is initially skeptical, and with the addition of psi positive studies, the hit rate increases to a high of 26.5% after the second study. With the addition of the remaining studies, the confidence interval shrinks, but the final hit rate of 26.3% still contains the chance hit rate of 25%. In contrast, the believer began with a 44.8% hit rate. With the addition of the first and second study, it does not deviate very far from the original value. With the addition of the remaining four studies the believer's estimate of the hit rate gradually

drops to 42.1% with its confidence interval never containing the chance hit rate of 25%.

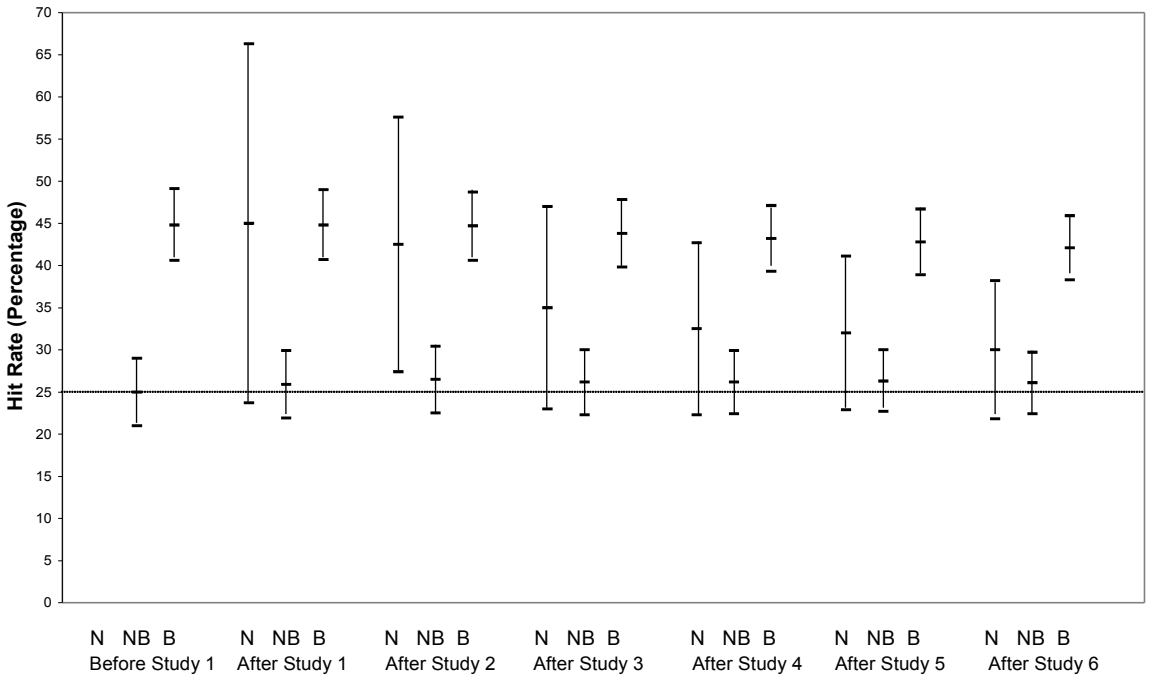

Figure 2. Mean hit rates and 95% confidence interval for Bayesian analysis using non-informative (N), non-believer (NB) and believer (B) prior probabilities. The dotted line represents the hit rate under the null hypothesis.

After six studies, the Bayesian analysis using a non-informative prior (equivalent to the meta-analysis) resulted in a failure to reject the null hypothesis. The non-believer also fails to reject the null hypothesis, whereas the believer rejects the null hypothesis. As with meta-analyses in general, with the addition of more studies the width of the confidence interval shrinks, representing the growth in confidence in the effect size estimate as more data is considered.

The results so far further demonstrate the improvement of the interpretation of findings from the Bayesian approach over the meta-analytic approach. If interpreted from a strict NHST perspective, the meta-analysis described earlier yields a non-significant finding, failing to provide evidence for psi phenomena. In other words, if the null hypothesis is true (that the magnitude of psi phenomena is zero), then

the likelihood of observing the data that we observed (or more extreme) is greater than 5%. Whereas this perspective requires the researcher to assume that the results of the data (30%) are conditional on the null hypothesis being true (p[data|hypothesis]), the Bayesian interpretation frames this from the perspective of the probability of a belief (or hypothesis) conditional on the observed data (p[hypothesis|data]). Therefore, researchers are encouraged to improve upon their hypothesis by collecting more and more data. From the perspective of an advancing science, the Bayesian perspective allows for and encourages the improvement of a hypothesis of the psi effect by increasing experimentation and accumulating evidence.

Bias assessment of the psi Ganzfeld literature

The extent to which any literature is impacted by publication bias is ultimately unknown. There are established methods to assess the extent of publication bias, such as using funnel plots (Begg, 1994) or the fail-safe n method (Rosenthal, 1979). At the same time however, these methods may be misleading (Macaskill, Walter & Irwig, 2001; Tang & Liu, 2000) or may yield vague results (e.g., does n studies in the file drawer seem likely for a particular area of research?). Despite the bias problem and difficulties associated with assessing the severity of the problem, current practice in psychology continues to rely on the meta-analysis of published literature to integrate research findings. One might argue that with the accumulation of *all* new findings, regardless of their results (hence there is no file drawer effect) one would be able to eventually undo any existing bias.

Let's assume we can begin to do this by using a set of data whereby we know there is no file drawer effect, that is, the data presented above. Let's also assume that the psi effect is in reality zero. If we proceed as if a cumulative science is self-correctable, we can examine the impact of adding the six studies to the existing literature. In Table 3, the Bayesian analysis of three recent meta-analytic results (Bem & Honorton, 1994; Milton & Wiseman, 1999; Storm & Ertel, 2001; see Appendix for summary of the meta-analyses) with the six current

studies is presented. In contrast to the previous analysis, rather than using illustrative non-informative, believer and non-believer priors, the meta-analytic effect sizes of the Bem and Honorton (1994), Milton and Wiseman (1999) and Storm and Ertel (2001) served as initial priors. The meta-analytic priors are similar to those presented earlier in that there is some discrepancy in their hit rate estimates.[5] The graphical representation of this analysis with 95% confidence intervals is depicted in Figure 3.

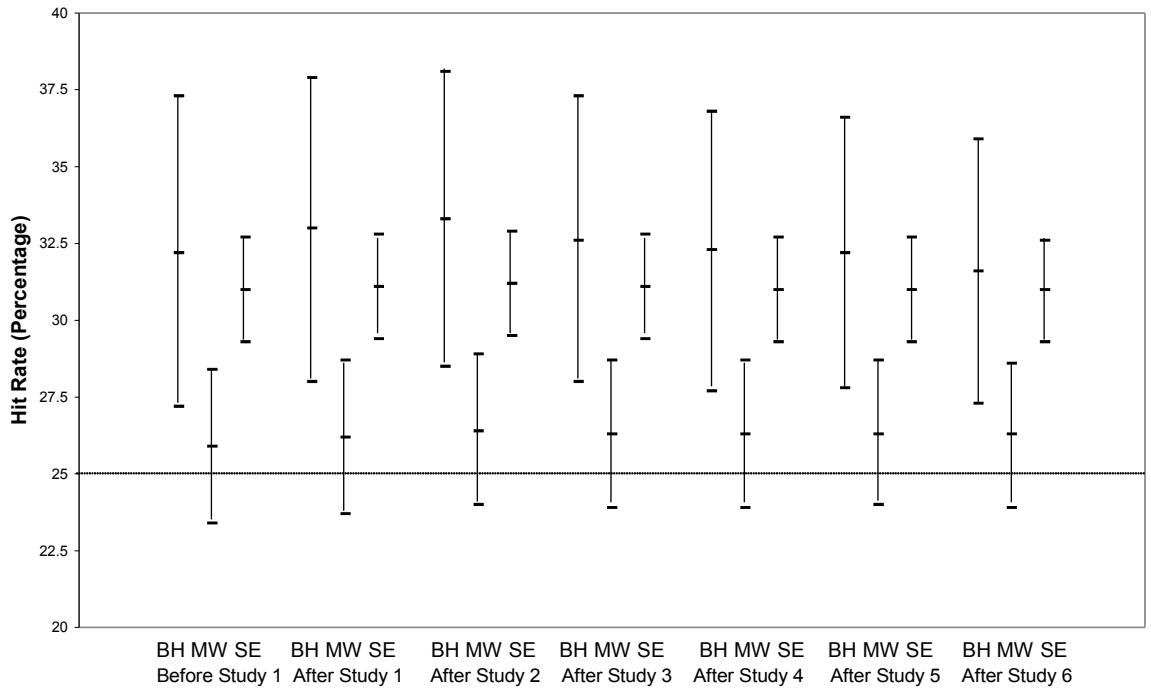

Figure 3. Mean hit rates with 95% confidence intervals for Bayesian analysis using Bem & Honorton (1994) (BH), Milton & Wiseman (1999) (MW) and Storm & Ertel (2001) (SE) meta-analyses as prior probabilities. The dotted line represents the hit rate under the null hypothesis.

[5] Although a majority of the meta-analyses examined used mean *ES* as the effect size measure for psi, we have chosen to use the hit rate instead. Not all psi Ganzfeld studies employ a four-choice judging set such that a 25% hit rate is equivalent to chance occurrence. Nevertheless, by expressing the meta-analytic findings in such a way, the reader can more intuitively understand the effect size than when using mean *ES*. The effect sizes expressed as hit rates were reported directly by Bem and Honorton (1994) and Storm and Ertel (2001). Together with the reported number of trials meta-analyzed, the information was used to derive a beta distribution prior using the method described in an earlier section. Milton and Wiseman (1999) did not directly report a hit rate for their meta-analysis. To create a consistent and intuitive effect size measure, a hit rate was derived by combining the reported *z* scores (unweighted Stouffer *z* method) for each of the studies analyzed by Milton and Wiseman and calculating the number of hits out of the total trials that would result in an exact binomial significance equivalent to that *z* score.

As with the previous analysis, priors with relatively wide CIs will be more impacted by the data. It is not difficult to see that this is a function of the sample size of the meta-analysis. When the sample size (or the number of total trials) of the meta-analysis is large, the variance of the mean estimate is invariably smaller than one with a smaller sample size. Bem and Honorton's (1994) meta-analysis contained the smallest number of trials (see Appendix) and hence with the addition of the six studies the CIs shrunk more so than the other two meta-analytic priors. The degree to which the point estimate of the prior will be impacted by the data is also in part a function of the variance of the prior.

In proceeding with our goal of "correcting" the literature or "undoing" the bias, we can see from Figure 3 that when a literature is very large, it becomes challenging to accomplish this. In our case with the hypothetical assumption that the psi effect is in reality zero, the effort to drive the two significant findings (BH and SE) to non-significance is even more hopeless. As more null studies are added, the CI is concurrently shrinking such that getting the CIs to eventually contain the null value requires a substantive number of null studies. This points to the dilemma of drawing conclusions from a large and biased literature and the difficulty that adding new unbiased data can eventually correct the bias.

Although it may be reasonable to make the assumption that by eliminating the file drawer problem, the six studies conducted will provide unequivocal support for or against a psi effect, the mere elimination of publication bias is not a panacea. This point is strengthened when we examine Bem and Honorton's (1994) meta-analysis. They reported that their analysis is also not plagued by the file drawer problem as they reported all trials associated with an autoganzfeld research program spanning 6 years. According to that meta-analysis, the effect size is significant with a point estimate of 32.2% (mean ES = .162, p = .002). Both Bem and Honorton (1994) and the six studies conducted here are free of the publication bias problem, yet their conclusions are not in agreement. The removal of the selective

publishing bias does not necessarily yield consistent and conclusive results.

Table 3. Mean hit rates and standard deviations for meta-analytic priors and posterior probabilities

	Bem & Honorton (1994)		Milton & Wiseman (1999)		Storm & Ertel (2001)	
	Hit rate	SD	Hit rate	SD	Hit rate	SD
Meta-analytic prior	32.2%	3	25.9%	1	31.0%	<1
After Study 1	33.0%	3	26.2%	1	31.1%	<1
After Study 2	33.3%	2	26.4%	1	31.2%	<1
After Study 3	32.6%	2	26.3%	1	31.1%	<1
After Study 4	32.3%	2	26.3%	1	31.0%	<1
After Study 5	32.2%	2	26.3%	1	31.0%	<1
After Study 6	31.6%	2	26.3%	1	31.0%	<1

Discussion

The overarching questions that have occupied parapsychological research has been, "Does psi exist?" and "To what extent have decades of psi research served to support or refute the existence of it?" We conducted a series of psi Ganzfeld trials and compared analyses of the results from NHST, meta-analytic and Bayesian approaches. The series of six studies conducted here is an attempt to contribute to this discussion both on a methodological and substantive level.

As many others have argued, the synthesis of research findings is important in the assessment of the evidence accumulated across years of psi Ganzfeld research (Krippner *et al*., 1993; Rosenthal, 1986; Utts, 1991). The meta-analytic treatment of the current data with three previously published meta-analyses is an attempt at synthesizing results across studies. Results indicate that with the addition of the six studies to the meta-analytic databases, the effect size estimates do not change drastically and that conclusions based on the rejection/acceptance of the null hypothesis also do not change as a result of the new findings. Storm and Ertel (2001) with the largest number of studies meta-analyzed saw their 31% estimate fail to change within a tenth of a percentage point with the addition of all six studies. Both Bem and Honorton (1994) and Milton and Wiseman's (1999) estimates changed within a sixth of a percentage point of the original estimate. The recent psi meta-analytic literature remains somewhat inconsistent with one negative and two positive estimates.

One of the criticisms of the meta-analytic literature has been the problem associated with biased estimates as a result of the file drawer problem (Hyman, 1985; Kupfersmid, 1988; Rosenthal, 1979). We hope to have demonstrated that the removal of this bias is not in and of itself a panacea. Ruling out the problem of selective reporting does not preclude the need for studies to be conducted in a procedurally and methodologically sound manner. The necessitation for studies to have adequate power (Cohen, 1992) and consistently demonstrate replication of effect sizes (Krippner *et al*., 1993; Rosenthal, 1986; Utts, 1991) are two suggestions for developing a coherent and convincing literature.

Howard and his colleagues (Howard *et al*., 2000, Howard, Hill *et al*., 2009) have argued that the overreliance on NHST in both primary statistical analyses and meta-analyses contributes to some of the inconsistencies and misinterpretation of research findings. We have tried to suggest as they have, that the Bayesian perspective provides an alternative framework upon which to analyze data. The research literature as it currently stands is inherently biased due to the file drawer problem. NHST is one of the reasons that this problem persists.

The Bayesian approach is an improvement over NHST framework in that it is both more intuitive and straightforward to interpret and it more effectively facilitates the synthesis of data in the estimation of an effect. One of the more controversial aspects of the approach is the explicit recognition of subjectivity in the analysis of data (Press & Tanur, 2001). Although Howard and his colleagues have pointed out that the subjectivity inherent in priors, after enough data, would ultimately and largely reflect empirical findings, the data presented here demonstrate that the degree to which a hypothesis (expressed as priors) will be driven by progressive collection of new data is strongly dependent on the strength of the priors themselves.

Despite these criticisms, the Bayesian approach is promising as a alternative data analytic framework, especially when implemented in the context of research registries. The idea of research registries has been most firmly implemented with research conducted in the medical field. A 2004 policy adopted by the International Committee of Medical Journal Editors (ICMJE, 2004; impacting journals such as *Journal of American Medical Association*, *The Lancet*, *New England Journal of Medicine*) required that all clinical trials be entered in a public registry before the onset of patient enrollment, as a condition of consideration for publication. Subsequently, the United States Food and Drug Administration (FDA) Amendment Act of 2007 further required all clinical trial registration of studies on drugs and devices under their jurisdiction on the http://www.clinicaltrials.gov/ website. With the establishment of a research registry, the problem related to biased reporting can be more effectively minimized. Moreover, groups of scientists can establish procedural and methodological guidelines to ensure the consistent quality of studies being conducted and avoid the task of having to fix or undo a flawed literature.

The answer to the substantive question of whether psi exists and whether years of psi Ganzfeld research provide evidence for or against it cannot be adequately provided without addressing the methodological issues raised in this study. Common problematic publication practices that contribute to the file drawer problem coupled with the dominance of NHST approaches in psychology

largely contribute to an inconsistent and confusing literature. Furthermore, under this scenario, the hope for a self-correcting science may be more of an illusion than reality. Consequently, this study and other researchers (Howard *et al*., 2000, Howard, Hill *et al*., 2009; Howson & Urbach, 1989; Kline, 2004; Press & Tanur, 2001) call for the consideration of alternative analytic approaches such as Bayesian methods. Although the Bayesian perspective does not dominate science, there is support and application of Bayesian methods in a number of fields including mainstream social sciences (e.g., American Psychological Association Task Force on Statistical Inference, 2000; Gill, 2002), economics (Lancaster, 2004), and medicine (Ashby & Smith, 2000). Parapsychology having been an area of science put under particularly strong scrutiny for defending its findings would benefit from considering such alternatives in which other mainstream areas of science have also realized the benefits of adopting.

Acknowledgements

This study was funded in part by a research grant from the Parapsychology Foundation, Inc. and was completed as part of Michael Y. Lau's master's thesis at University of Notre Dame. It was incompletely presented and referenced in Delgado-Romero & Howard (2005), and Howard, Lau *et al.* (2009).

References

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*, 119-126.

Ashby, D., & Smith, A. F. M. (2000). Evidence-based medicine as Bayesian decision-making. *Statistics in Medicine*, *19*, 3291-3305.

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

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

Blackmore, S. (1980). The extent of selective reporting of ESP Ganzfeld studies. *European Journal of Parapsychology*, *3*, 213-219.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.). *The Handbook of Research Synthesis*, (pp. 193-213). New York: Russell Sage Foundation.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

Dalkvist, J. (2001). The Ganzfeld method: Its current status. *European Journal of Parapsychology*, *16*, 19-22.

Delgado-Romero, E. A., & Howard, G. S. (2005). Finding and correcting flawed research literatures. *The Humanistic Psychologist*, *33*, 293-303.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology*, *5*, 75-98.

Folger, R. (1989). Significance tests and the duplicity of binary decisions. *Psychological Bulletin*, *106*, 155-160.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Karen & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311-339). Hillsdale, NJ: Lawrence Erlbaum.

Gill, J. (2002). *Bayesian methods for social and behavioral sciences*. Boca Raton, FL: Chapman & Hall.

Hardy, A., Harvie, P., & Koestler, A. (1973). *The challenge of chance*. New York: Random House.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, *88*, 359-369.

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

Honorton, C. (1997). The Ganzfeld novice: Four predictors of initial ESP performance. *The Journal of Parapsychology*, *61*, 143-158.

Honorton, C., & Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *Journal of the American Society for Psychical Research*, *68*, 156-168.

Howard, G. S., Hill, T. L., Maxwell, S. E., Baptista, T. M., Farias, M. H., Coelho, C., Coulter-Kern, M., Coulter-Kern, R. (2009). What's wrong with research literatures? And how to make them right. *Review of General Psychology*, *13*, 146-166.

Howard, G. S., Lau, M. Y., Maxwell, S. E., Venter, A., Lundy, R. R., & Sweeny, R. M. (2009). Do research literatures give correct answers? *Review of General Psychology*, *13*, 116-121.

Howard, G. S., Maxwell, S. E., & Fleming, K. J. (2000). The proof of the pudding: An illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis. *Psychological Methods*, *5*, 315-332.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle, IL: Open Court.

Hubbard, R. Parsa, R. A., & Luthy, M. R. (1997). The spread of statistical significance testing in psychology. *Theory and Psychology*, *7*, 545-554.

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

International Committee of Medical Journal Editors (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. Retrieved April 29th, 2009 from http://www.icmje.org/clin_trial.pdf

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, D.C.: APA.

Krippner, S., Braud, W., Child, I. L., Palmer, J., Rao, K. R., Schlitz, M., White, R. A., & Utts, J. (1993). Demonstration research and meta-analysis in parapsychology. *Journal of Parapsychology*, *57*, 275-286.

Kupfersmid, J. (1988). Improving what is published: A model in search of an editor. *American Psychologist*, *43*, 635-642.

Lancaster, T. (2004). *An introduction to modern Bayesian econometrics*. Malden, MA: Blackwell.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, *20*, 641-654.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of 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.

Oakes, M. (1986). *Statistical inference: A commentary for the social behavioral sciences*. New York: John Wiley & Sons.

Pollard, P. (1993). How significant is "significance"? In G. Karen & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 449-460). Hillsdale, NJ: Lawrence Erlbaum.

Press, S. J., & Tanur, J. M. (2001). *The subjectivity of scientists and the Bayesian approach*. New York: Wiley-Interscience.

Pruzek, R. M. (1997). An introduction to Bayesian inference and its applications. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 287-318). Mahwah, NJ: Lawrence Erlbaum.

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

Rosenthal, R. (1986). Meta-analytic procedures and the nature of replication: The Ganzfeld debate. *Journal of Parapsychology*, *50*, 315-336.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.). *The Handbook of Research Synthesis*, (pp. 231-244). New York: Russell Sage Foundation.

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

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173-1181.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115-129.

Storm, L., & Ertel, S. (2001). Does psi exist? Comments on Milton and Wiseman's (1999) meta-anlysis of Ganzfeld research. *Psychological Bulletin*, *127*, 424-433.

Tang, J., & Liu, J. L. Y. (2000). Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology*, *53*, 477-484.

Task Force on Statistical Inference (2000). Narrow and shallow. *American Psychologist*, *55,* 965-966.

Utts, J. (1986). The Ganzfeld debate: A statistician's perspective. *Journal of Parapsychology*, *50*, 393-402.

Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science*, *6*, 363-403.

Appendix: Summary of Three Published Psi Ganzfeld Meta-Analyses

Meta-Analysis	No. of studies	No. of Trials	Mean *ES*	Stouffer *Z*	*p*
Bem & Honorton (1994)	10	329	.162	2.52	.002
Milton & Wiseman (1999)	30	1198	.013	.699	.24
Storm & Ertel (2001)	79	2767	.138	5.66	7.78×10^{-9}

Note: The three meta-analyses do not contain independent trials; for example, Storm and Ertel (2001) included studies meta-analyzed in both Bem and Honorton (1994) and Milton and Wiseman (1999).

European Journal of Parapsychology
Volume 24.1, pages 32-52

ISSN: 0168-7263

Staring Nowhere? Unseen Gazes Remain Undetected under Consideration of Three Statistical Methods

Thomas Wehr
University of Trier, Germany

Abstract

With a direct-looking experiment, the remote staring effect was investigated. Two people sat in two rooms separated by a one-way mirror. In some trials the experimenter stared at the participant who tried to detect the staring signal. One group was aware of the number of staring and non-staring trials and another group did not know how many trials would follow. In addition to the sign method according to Sheldrake (2005) and to the ratio calculation according to Schmidt (2005), signal detection theory was conducted on the staring data to take response bias into account. An overall conservative response bias was evident and mean sensitivity was close to zero. Awareness of the number of staring and non-staring trials had no effect on sensitivity and response bias. The results are discussed in light of the current methodologies and suggestions are made about alternative research strategies.

Introduction

Did you ever feel a gaze from behind *before* you recognised that someone is *really* staring at your back? The phenomenon of remote staring detection is discussed in the scientific literature since the nineteenth century (Titchener, 1898; see Braud, Shafer, & Andrews,

Correspondence details: Dr. Thomas Wehr, Department of Psychology, University of Trier D-54286, Trier, Germany. Email: wehr@fishpost.de.

1993; Colwell, Schröder, & Sladen, 2000; Schmidt, Schneider, Utts, & Walach, 2004; Sheldrake, 2005, for literature reviews). The term "remote staring" refers to the experience of sensing an unseen gaze without seeing the gazing person.

Two lines of research with two different categories of dependent variables can be identified. First, in *direct-looking experiments*, a participant is stared at by another person who usually sits behind the participant. The spatial setting can be composed of one room with close proximity between starer and staree (e.g., Sheldrake, 2003), or both individuals are separated by a one-way mirror (e.g., Colwell et al., 2000). With the latter setting, important confounding variables can be eliminated (e.g., sensory leakage). The success of direct-looking experiments is most commonly measured by the *direct response method*. The staree gives an intended response whether he is stared at or not. Second, in *indirect-looking experiments*, closed circuit television systems are used, with the staree and the starer in separate rooms (EDA-CCTV; e.g., Braud et al., 1993). The starer does not directly focus on the back of the staree, but rather looks at a monitor that is connected with the staree. The success of indirect-looking experiments is most commonly measured by indirect response parameters such as physiological variables (e.g., electrodermal activity). The present study describes a direct-looking experiment and the success of staring was measured by direct responses, that is, the experimenter directly stared at the participant from behind while the participant was aware about the intention of the experiment and made an attempt to detect the unseen gazes.

Does the remote staring effect really exist? Depending on the experimental setting and the response method, the empirical findings are far from clear-cut. Even conclusions drawn from meta-analyses do not definitely verify or falsify the existence of the remote staring effect. For example, Schmidt et al. (2004) reported a small but significant effect across fifteen EDA-CCTV experiments. However, methodological quality of the selected studies – rated by experts – was only 61%. Thus, it is not clear, whether the reported effects are artefacts

or evolved from an existing sense of being stared at. Consequently, the authors advocated high-quality replication studies.

Another meta-analysis (Radin, 2005) was conducted on sixty studies, including several studies from Rupert Sheldrake. Radin differentiated high and low quality studies as well and observed lower effect sizes in high quality studies compared to low quality studies, but the effect was still significant even in the high quality studies. However, a methodological flaw in his meta-analysis can be identified in the calculation of the dependent variable: A ratio was calculated by dividing a correct yes-response (hit) by the total number of trials. Thereby, response bias was not taken into account. However, in a two-alternative forced choice task (i.e., the staring experiment with the direct response method) participants do not naturally adopt a neutral response strategy (i.e., they say "yes" equally often as they say "no"). It is rather reasonable to assume that participants accommodate their response strategy to their pre-assumptions. For example, if a participant is suspicious about the existence of the staring effect, saying "no, I was not stared at" may be the more appropriate response strategy than saying "yes, I was stared at". It is also possible that participants want to be successful and show a bias towards saying "yes". Such response strategy would increase the probability to detect an unseen gaze because during the experiment a positive staring trial is expected to occur sooner or later. Most importantly, the neglect of such response bias is the kernel of criticism about the Sheldrake studies (see Atkinson, 2005; Schmidt, 2005).

Sheldrake (2005) used a "sign" test which divides the sample into one group with participants who are more correct than incorrect (+) and another group with participants who are more incorrect than correct (-). Participants who are equally right and wrong are excluded. Both groups can then be compared by a chi-square test. The sign method must be criticised for at least two reasons. First, downsizing the sample reduces test power. For example, Sheldrake (1999, Table 2) collected 1.744 trials. After division into a plus and a minus group, only 146 observations remained for the statistical test. Under consideration of a commonly reported small effect ($\omega^2 = .02$) the test

power would be around chance ($1 - \beta = .50$) instead of $1 - \beta > .99$ with all 1.744 trials. Second, the sign method reduces the phenomenon itself because all participants are equally weighted. In other words, a person who detects ten of ten unseen gazes (i.e., 100% ability in detecting unseen gazes) and correctly says "no" on all non-staring trials would be treated in the same way as a person who detects only one of ten unseen gazes (i.e., 10% ability in detecting unseen gazes) and correctly says "no" on all non-staring trials. Moreover, participants who are equally right and wrong, will be excluded from the sample. But why that? Data exclusion for mere statistical reasons won't help explain a poorly understood phenomenon like the remote staring effect. However, the sign method was conducted on the present data to provide comparability with Sheldrake's results.

Schmidt (2005) argued that the data presentation of Sheldrake (e.g., 2005, Figure 1A; Figure 4 in the present study) can lead to wrong conclusions because staring and non-staring trials are presented separately without taking response bias into account. Sheldrake concluded from his results that people are better in detecting unseen gazes on staring trials than to detect the absence of a gaze on non-staring trials. However, a response bias may be responsible for this. Therefore, Schmidt (2005) recommended to divide correct yes-responses by overall yes-responses and to divide correct no-responses by overall no-responses. In doing so, Schmidt found no difference in the ability to perform correctly on staring and non-staring trials in Sheldrake's data.

In conclusion, the ratio method is clearly more valid than the sign method. The sign method harbours the risk of low test power and disregards intrapersonal as well as interpersonal differences in response bias, thus blurring some important characteristics of the phenomenon. In contrast, the ratio method considers the basic response-rates of yes- and no-responses, thereby reducing the impact of response bias and thus allocating data which capture the phenomenon more accurately.

A third possibility for statistical treatment of the staring data was proposed by Atkinson (2005). In direct-looking experiments using the

direct response method the question is always whether participants detect the staring signal sent by the starer or not, that is, how sensitive their ability is to sense unseen gazes. Thus, a measure is needed that represents sensitivity. Moreover, the measure should be valid to adjust for response bias. Finally, a measure would be helpful that makes response bias visible. All three requirements are fulfilled by the appliance of signal detection theory (SDT; Macmillan & Creelman, 2005; McNicol, 2005). SDT was initially developed in the field of psychophysics to interpret sensory processes (McNicol, 2005). The relevance of signal detection theory to psychology lies in the fact that it is a theory about the ways in which choices are made, especially, about the ways in which people discriminate between presence and absence of a certain perceptual cue (called "signal"; in the present study it is the staring signal) within an indifferent background (called "noise"; i.e. any other present sensory information). Typically, a yes-no task with plenty of trials is applied to determine the sensitivity of a participant (i.e., the ability to identify the signal). On half the occasions the signal is absent and only noise is present (i.e., non-staring trials). On other occasions noise plus a signal is shown (e.g., a staring signal on staring trials). Noise, and signal plus noise trials occur at random. After each trial the participant must say whether it was a signal plus noise trial or just noise alone.

Applied to a direct-looking experiment, there are four possible outcomes on each trial: On staring trials participants produce either *hits* (i.e., correctly saying "yes") or *misses* (i.e., missing the signal and saying "no"). On non-staring trials participants produce either *correct rejections* (i.e., correctly saying "no") or *false alarms* (i.e., mistakenly saying "yes", though no signal is present).

Sensitivity refers to the ability to discriminate between the presence and the absence of a staring signal. The most common sensitivity measure in SDT is called d' which is defined by hits minus false alarms, both converted into z scores: $d' = z(Hits) - z(False\ Alarms)$. A $d' = 0$ indicates no accuracy at all (hits = false alarms), perfect accuracy implies an infinite d'. Therefore, if the remote staring effect exists, d' must significantly differ from zero in a positive direction.

Atkinson (2005) calculated a *d'* for staring data obtained from Sheldrake (1999, Table 5) and found a *d'* = .25. However, he failed the option to test whether sensitivity was significantly different from zero.

In addition, response bias can be defined as $c = -0.5\ [z(Hits) + z(False\ Alarms)]$. A $c = 0$ indicates the absence of any response bias, that is, when the false alarms and missing rates are equal. With a response bias towards saying "no", *c* becomes a positive value (conservative response criterion). On the other hand, with a response bias towards saying "yes", *c* becomes a negative value (liberal response criterion). A conservative response criterion is characterised by a low proportion of hits and false alarms and a high proportion of correct rejections and misses, whereas a liberal response criterion is characterised by a high proportion of hits and false alarms and a low proportion of correct rejections and misses.

In conclusion, the sensitivity index *d'* seems to be the most adequate statistical index to calculate staring data using the direct-looking method. However, the appliance of *d'* in published articles is rare (e.g., Atkinson, 2005). At least, some attempts were made to adjust for response bias. For example, Radin (2004) adjusted data from many Sheldrake studies and two older studies (Coover, 1913; Poortman, 1959) for response bias. His results were consistent with the original results. However, Radin admits that "the design did not absolutely exclude subliminal sensory cues" (p. 250). Thus, because of flaws in the selected studies of Sheldrake (i.e., sensory leakage), the adjusted data are not very helpful.

In the present study, new staring data were collected and calculated in three different ways: (1) sensitivity and response bias according to SDT, (2) the sign method according to Sheldrake, and (3) the ratio calculation according to Schmidt. participants underwent a direct-looking experiment with ten staring and ten non-staring trials in truly randomised order. It was argued that the ability to detect unseen gazes has to be separated from individual tendencies to respond with "yes" or "no" (response bias). The SDT-method should be sufficient to identify response bias and to detect the staring effect, if it really exists. In addition, one group received detailed information about the number

of staring and non-staring trials, whereas an uninformed group did not know how many trials would follow. This was done to test whether such information influences response bias, that is, in the present study, knowing the number of staring and non-staring trials should lead to a neutral response criterion whereas not being informed about the staring/non-staring ratio should make the participant more cautious about saying "yes" (conservative response bias).

Method

Participants and Design

Sixty-four students of the University of Trier participated in the experiment (48 females and 16 males; mean age: 22 years). Forty-nine participants were undergraduate students of psychology in the first academic year, whereas fifteen participants were enrolled in other fields of study (i.e., geography, and pedagogy). All participants were rewarded for their time with course credits.

The sample was divided into a full instruction group and an uninformed group: Thirty participants received detailed information about the number of staring and non-staring trials (i.e., the full instruction group: "Twenty trials with ten staring trials and ten non-staring trials in randomised order"), whereas thirty-four participants were not informed (i.e., the uninformed group: "On some trials someone will stare at you and on other trials no one stares"). Assignment to one of the two instruction groups was randomised. All participants underwent twenty trials with ten staring trials and ten non-staring trials. To avoid implicit sequence learning, no feedback was given and the staring/non-staring trial order was truly randomised.

Materials

The experiment was a double-blind study, that is, participants were given no feedback about their rating accuracy and the

experimenter received his staring/non-staring instructions only trial-by-trial via the computer. Thus, expectation effects as well as experimenter effects were widely diminished.

In detail, for each participant, a truly randomised sequence order was generated during the ongoing experiment.[1] For this purpose a program (Pepperoni v0.1) was written that automatically generates and presents genuine random numbers trial-by-trial. The numbers (ones and zeros) were taken from an online resource.[2] The 20-trial sequence was then presented step-by-step, that is, only one number at the moment was visible (1 = staring, 0 = non-staring). At the same time, only the number of trials so far passed and the number of trials to come were visible. Then, the experimenter turned to the next trial by pressing a button (see Figure 1).

A second program was written for the participants who had to decide at the end of a trial by clicking on one of two fields whether they had felt stared at or not. In addition, participants received audio signals at the beginning of the inter-trial pause, at the beginning and at the end of a test trial. During a test trial the display remained black. The inter-trial pause lasted 5 seconds and was adapted to provide enough time for the starer to proceed to the next staring/non-staring order in Pepperoni v0.1, while participants had time to relax. A test trial lasted 10 seconds and closed with an audio signal and the request to evaluate the experience (i.e., stared at or not stared at). For this decision, participants had unlimited time. After response, the next inter-trial pause started immediately (see Figure 1).

[1] Randomisation sequence is available from the author on request.

[2] This internet resource was HotBits (www.fourmilab.ch/hotbits/) that uses a commercial Geiger-Müller detector with a 5 microcurie Cæsium-137 check source to derive genuine random numbers from radio decay (for further details, see http://www.fourmilab.ch/hotbits/hardware3.html).

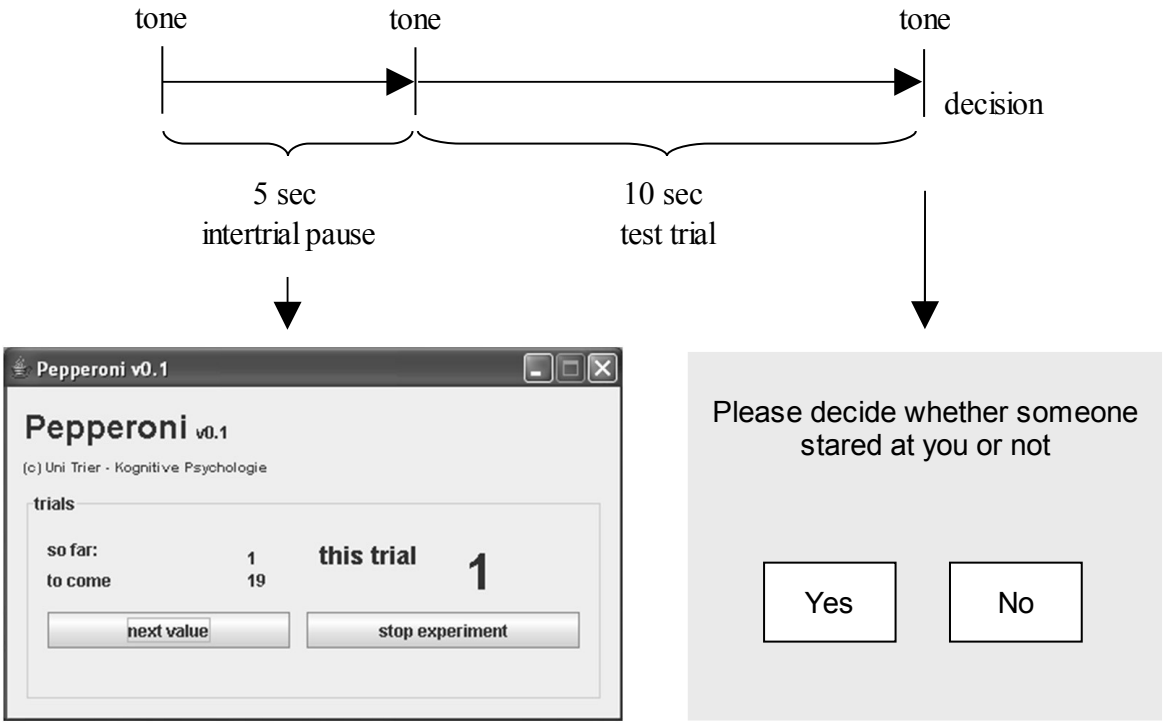

Figure 1. A schematic description of the procedure. After a tone, the 5 sec-inter-trial pause preceded the 10 sec-test trial. In the meantime, the experimenter clicked on the "next value" button of the randomization program and received the staring/non-staring order for the next test trial. A second tone started the test trial and the experimenter sent the staring signal or not, while the participant tried to detect the signal. A third tone stopped the test trial and participants were asked to decide whether they were stared at or not. This procedure continued until all twenty test trials were performed.

It was assumed that the staring effect is rather small and that such a weak signal is easily masked by internal somatic "noise". By reducing sensory input, person's ability to detect the staring signal should be enhanced (see Bem, & Honorton, 1994; Honorton, 1977, for comparable interventions with related *psi*-phenomena). To reduce internal signals from the body in the present study, a relaxation exercise was conducted at the beginning of the experiment. Together with the experimenter, all persons underwent a five-step breathing exercise from Krampen (2004).

The experimental setting can be seen in Figure 2. Two rooms were divided by a one-way mirror. All participants sat in front of a computer and with the back to the window in a well-lit room. The experimenter's room was not illuminated and the experimenter wore

dark clothes to minimise the risk of movement detection. The experimenter sat at the back wall of the room and ran the Pepperoni v0.1 program via a notebook. The brightness of the display was minimised and covered by the cabinet door. In non-staring trials, the cabinet door entirely blocked the line-of-sight of the starer to minimise the risk of unwanted staring. In staring trials, the experimenter slightly bent to the left to stare at the participant. An audio system from a television was used to transfer the audio signals from the participant's computer to the experimenter's room.

Procedure

Two of five experimenters worked always together. Experimenter 1 supervised the participant, whereas the second experimenter was never seen by the participant. Actually, he stayed in the experimenter's room and solely acted as starer. The author was open-minded, but never stared or acted as experimenter. The five assistants were all female, 22 years old and students of the 2nd year. They acted as starer or experimenter to a nearly comparable amount. All five assistants were open-minded and curious about their own staring abilities.

Figure 2. Experimental setting with two rooms separated by a one-way mirror. The participant sat in a well-lit room with the back to the window (left), whereas the experimenter sat in a dark room and sent staring signals or not, guided by a randomization program (right).

Experimenter 1 welcomed the participant and directed him into the first room. The remote staring phenomenon was briefly described and the experiment was declared as a scientific attempt to observe the effect in the laboratory. The participant was also informed that a second person sat in the adjacent room and would sometimes stare at the participant and sometimes not. In the full instruction group, participants were informed that the experiment consisted of twenty trials with ten staring trials and ten non-staring trials in randomised order. The experimenter asked the participants to close their eyes when the tone signal announced the beginning of a trial and then to concentrate on any internal and external sensations and signals which could indicate a gaze from behind.

The relaxation exercises followed the instructions and lasted about five minutes. The intention was to achieve cognitive and physiological relaxation. Experimenter and participant sat comfortably and began with short tensions and long lasting relaxations of the shoulder muscles, followed by breathing exercises. Subsequently, the participant performed three practise trials without data collection. Thus, the participant was given the opportunity to become familiar with the procedure and to find out how remote stares can be detected. When the participant had no further questions, the first experimenter left the room and the second experimenter started both programmes simultaneously. The starer was instructed to stare at the neck and the back of the participant's head on staring trials. In addition, he thought something like "turn around" and "I stare at you". On non-staring trials he was instructed to turn away from the participant and think about something else. Finally, in addition to age, sex and the field of study, participants were asked about their degree of motivation, how exhausting it was, and whether they believed that the remote staring effect exists.

Results

Most of the participants said that they felt comfortable about the relaxation exercise at the beginning. After the experiment, 92% said

that they were motivated and 90% were not much exhausted by the procedure.

The results are reported in four steps. First, sensitivity and response bias were analysed. Second, the sign method was conducted on the data. Third, ratios were calculated by dividing correct yes-responses by all yes-responses. Finally, control analyses were conducted.

Sensitivity and response bias (SDT)

Overall, mean sensitivity was d' = .0464 (see Figure 3a). This was not better than null sensitivity, $t_{(63)}$ = 0.62, ns. Overall, the response criterion was c = 0.1342 (see Figure 3b). This was significantly above zero, $t_{(63)}$ = 3.35, p < .001, ω^2 = .14. Thus, participants were generally cautious and showed a bias towards saying "no".

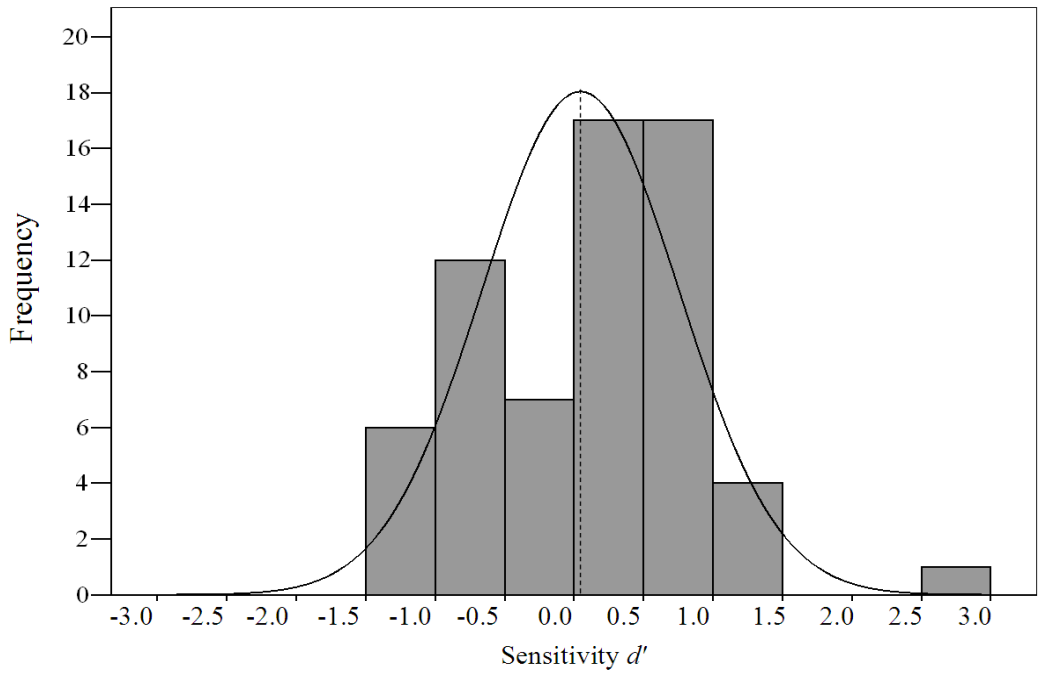

Figure 3a. Sensitivity (d') as a function of number of participants (frequencies), normal distribution curve and mean (dashed line) included.

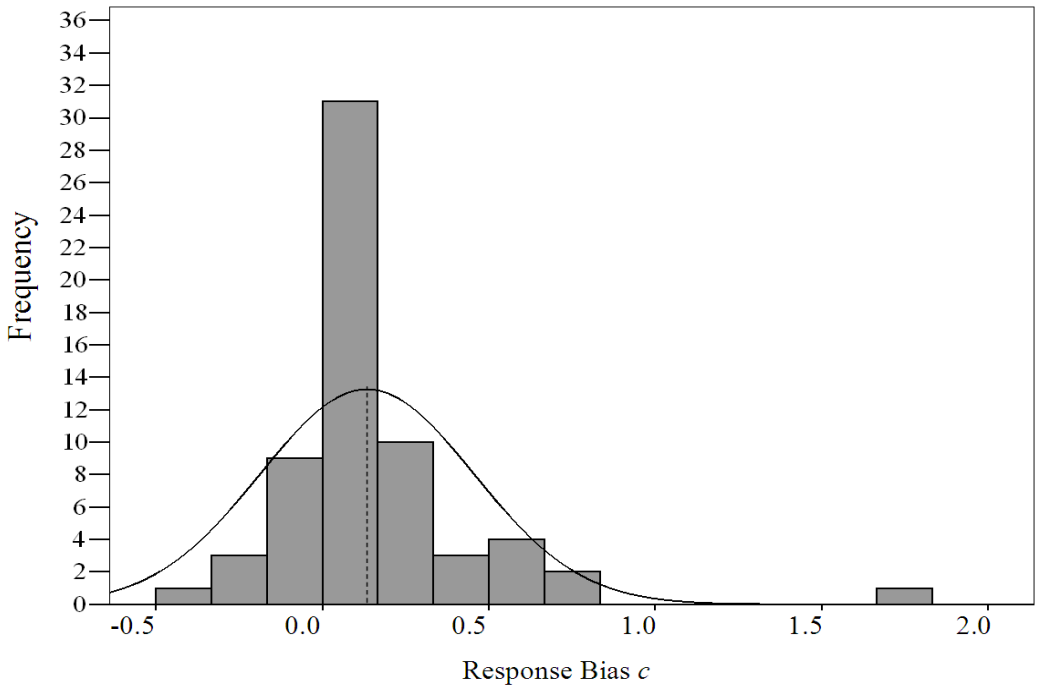

Figure 3b. Response bias (c) as a function of number of participants (frequencies), normal distribution curve and mean (dashed line) included.

In the full instruction group (d' = 0.0575) and in the uninformed group (d' = 0.0366), sensitivity was equal, $t_{(62)}$ = 0.12, *ns*. Moreover, both groups responded with a comparably conservative response bias, $t_{(62)}$ = –1.01, *ns*, although from a descriptive point of view, the uninformed group (c = 0.1723) was slightly more conservative than the full instruction group (c = 0.0912).

The sign method (according to Sheldrake)

As can be drawn from Figure 4, participants responded more often incorrectly than correctly on staring trials. In contrast, on non-staring trials, participants were able to give a considerable amount of correct rejections (54.7%) and made the fewest mistakes. Overall correct responses (50.3%) did only marginally exceed the chance level of 50%.

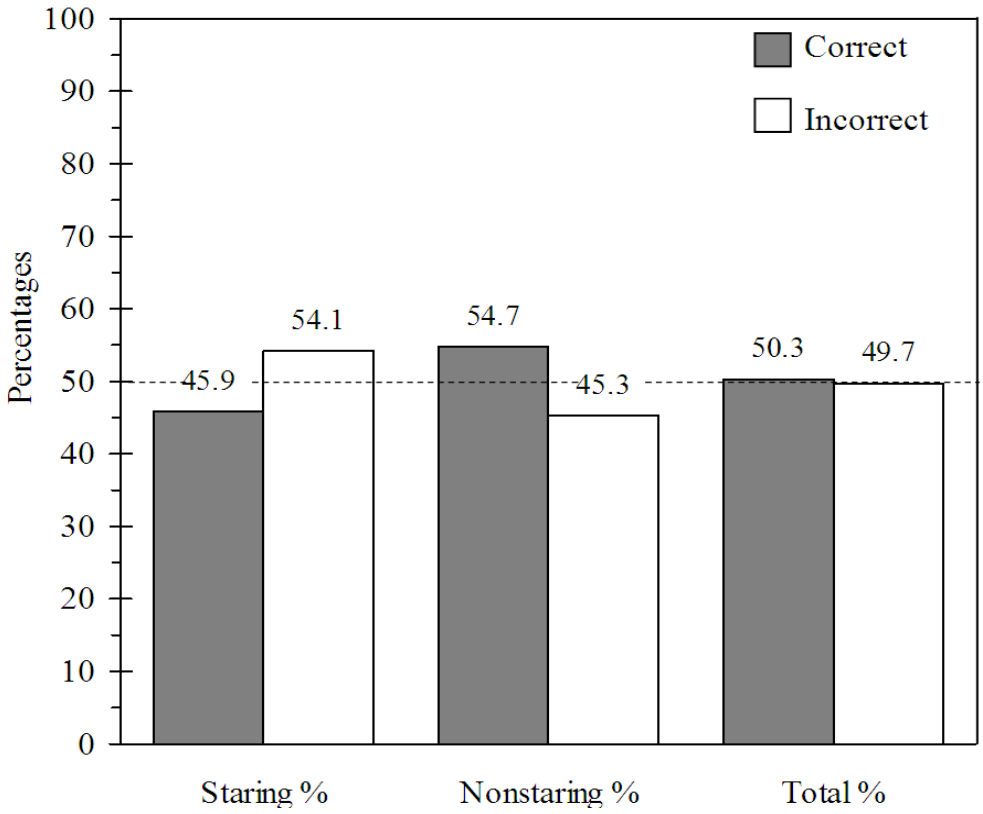

Figure 4. Mean proportions of correct and incorrect responses in staring trials, non-staring trials and in total (dashed line = chance level).

The data were transformed by Sheldrake's sign method (Sheldrake, 1999, 2005). The results are documented in Table 1. In contrast to Sheldrake (2005), whose results were "astronomically significant statistically" (p. 16), the chi-square test of the present experiment did not even reach an error probability of 10%. In other words, a staring effect could not be detected by the sign method here, neither for the complete sample nor for the subgroups divided by the instruction factor.

Table 1: Response frequencies, expressed as percentages of correct responses (chance level = 50 %) and in terms of signs, together with chi-square values and *p* values (cf. Table 1 in Sheldrake, 2005a, p. 15).

Sample	Trials	Correct	% correct	+	-	χ^2	*p*
Full Instruction Group	600	306	51.0	16	11	0.926	0.336
Uninformed Group	680	338	49.7	13	14	0.037	0.847
Total	1280	644	50.3	29	25	0.296	0.586

Note. + = number of participants who were more correct than incorrect; - = number of participants who were more incorrect than correct

Ratios (according to Schmidt)

participants made correct yes-responses (hits) in 51.3% of all yes-responses (hits and false alarms). Very similar, correct no-responses (correct rejections) were made in 49.9% of all no-responses (correct rejections and misses). Two one-sample *t*-tests revealed that both ratios did not differ from chance, $t_{(63)} = 0.70$, *ns*, and $t_{(63)} = 1.47$, *ns*. Moreover, both ratios did not differ from each other, $t_{(63)} = 1.43$, *ns*.

Control analyses

To test whether experimenter effects were present, control analyses were conducted. It was expected that the assistants would produce comparable staring data. Considering the staring sensitivity *d'*, this was the fact for all assistants in the role of the experimenter ($F_{(4,59)} = 1.41$, *ns*) and in the role of the starer ($F_{(4,59)} < 1$, *ns*).

Altogether, the results did not support the idea of a remote staring effect in the present sample. In contrast to Sheldrake (2005), no staring effect was detected with the sign method here. As can be drawn from Figure 4, participants of the present sample performed poorly on hits but produced a lot of correct rejections. Furthermore, they had only a small number of false alarms and a high missing rate. This data structure fits perfectly well with SDT. Overall response bias was conservative, that is, participants were very cautious in saying "yes", thereby missing the opportunity to produce a considerable amount of hits. Instead of that, the response bias towards saying "no" brought about high amount of correct rejections. Thus, even for participants with a sensitivity above chance (primarily the male participants), *d'* was mainly determined by correctly saying "no" on non-staring trials rather than by saying "yes", when in fact the experimenter stared at the participant.[3] However, this is not what the staring effect is about.

[3] The best male participant achieved a very high $d' = 2.57$. In particular, he made thirteen correct responses. However, this excellent performance was mainly the consequence of an extremely conservative response criterion ($c = 1.81$). In other words, the best male participant of the sample (participant ID 52, 20 years old, geography, believes in the staring effect) demonstrated an extraordinary sensitivity that was built on seventeen no-responses and only three yes-responses. All three yes-responses were hits (staring correctly identified; ratio = 100%) and he had ten correct rejections on non-staring trials (non-staring correctly identified; ratio = 77%). Thus, although he responded very cautious, he was able to detect staring signals and to reject them on non-staring trials with high reliability.

Moreover, it contradicts Sheldrake's observation that his participants guess at random on non-staring trials. In the present experiment, with 54.7% correct rejections, participants were right on non-staring trials significantly above chance, $t_{(63)} = 2.27$, $p < .03$, $\omega^2 = .08$. Sheldrake (2005) defines the detection of the absence of a stare an "unnatural request with no parallel in real-life conditions" (p. 18). However, participants in the present experiment did so successfully, so after all, such a request has a parallel in laboratory conditions.

Discussion

In the present study an attempt was made to detect the remote staring effect. Three different ways of statistical treatment with different quality were conducted on the data. Neither the sensitivity index *d'*, nor the sign method, nor the ratio calculation attested to the staring effect. Since the present study was conducted by five assistants, experimenter effects could be responsible for the null effect. Staring abilities can be sometimes diminished by experimenter effects because of different beliefs and opinions of the experimenters (Wiseman & Schlitz, 1997, 1999; Schlitz, Wiseman, Watt, & Radin, 2006). However, in the present study, no such experimenter effects were observed.

Surprisingly, a liberal response criterion was evident in Sheldrake's data (e.g., Radin, 2004), whereas in the present study, participants responded very conservatively. This shows that people behaved differently depending on the experimental setting (i.e., Sheldrake adapted a lot of field experiments). Moreover, *different* people reacted differently, that is, in the present study, the students clearly expected that unseen gazes can not be detected or they feared to make too many mistakes. participants were generally suspicious and very cautious in responding positively (i.e., saying "yes") to the task. The dominating presence of an overall conservative response bias made it difficult to detect the staring effect, because participants seldom said "yes", thus diminishing the probability of a hit. In the experimental situation, the participant's behaviour can largely depend on scepticism and pre-assumptions. For example, students of

psychology were more suspicious than students of geography and pedagogy, that is, 57% of the former said that they believe in the remote staring effect, whereas 64% of other fields of study agreed with the existence of the effect. Students of psychology demonstrated an almost perfect null sensitivity, whereas students of geography and pedagogy reached a d' of .22, which was very similar to the reported sensitivity by Atkinson (2005; $d' = .25$).

Another reason for a conservative response bias could have been the participant's attempt to work on the task very seriously and conscientiously. All students experienced the experiment as a welcome change in their regular experimental courses. Overall, they were suspicious but on the other hand, they were curious and enquiring, too. Thus, participants were willing to do the task with utmost care. Thereby, a conservative response strategy was efficient to minimise the risk of mistakes but also – as a side effect – the opportunity to produce a hit.

The conservative response bias can account for the low hit rate (the hit rate was below chance), but what does that mean for the existence of the staring effect? participants with a neutral response criterion (N = 19) produced a mean sensitivity of $d' = .0248$, which is not significantly different from zero, $t_{(18)} = 0.18$, *ns*. participants with a liberal response criterion and the highest hit rate (N = 13) did not even score above zero ($d' = -.1295$). Thus, the staring effect did not depend on any kind of response bias.

The results conflict with Sheldrake's data in three ways. In the present experiment, the sign method did not show evidence for a staring effect, a conservative response bias was found instead of a liberal response bias and performance on non-staring trials was not inferior to performance on staring trials.

At least in small samples, the sign method is an invalid statistical method to capture such a subtle phenomenon like the remote staring effect. If all trials are downsized to two categories, too much data loss drastically reduces the chance to detect the effect.

Sheldrake (2005) reported that people are better in detecting unseen gazes on staring trials than to detect the absence of a gaze on

non-staring trials. He argued that this data pattern is due to an evolutionary benefit for predator-prey relations. However, Schmidt (2005) demonstrated that response bias was responsible for this. In the present study, Schmidt's position was supported, that is, considering the ratios, successful performance on staring and non-staring trials was equal. Therefore, the idea of an evolutionary benefit seems unlikely.

The present lack of a staring effect is consistent with other findings (e.g., Lobach, & Bierman, 2004; Müller, Schmidt, & Walach, 2006). On the other hand, studies with a positive outcome are often of poor methodological quality (i.e., sensory leakage, implicit learning of sequences etc.). Sometimes it seems as if poor quality was covered by extremely large samples. It is a well known statistical principle that the probability of significance increased as the sample size increased. In other words, the smallest difference in means or the detection of the smallest signal will become significant by simply enlarging the sample. Thus, the absence of a staring effect in the present study can be due to a small sample size. However, even with extremely large samples, the data can erroneously produce a staring effect, too. For example, Schmidt (2005) pointed out that the staring effect was successively reduced by eliminating confounding variables such as direct feedback sensory cues. Of course, the staring effect seems to be very small, if it really exists. However, simply enlarging the sample size can not be the solution.

Staring studies had to continue in improving its quality. With the present study, a first small step forward was done by conducting SDT on staring data. However, further improvements are recommended.

First, response bias seems to be an important source of variance in staring experiments. Therefore, response bias should be conducted as an experimental variable. This can easily be done by the use of different pay-off matrices or via instructions, that is, telling people the (fictive) ratio of staring to non-staring trials (for further details, see Macmillan & Creelman, 2005; an alternative opportunity is discussed by Schmidt, 2005).

Second, in specific samples it may be helpful to increase the plausibility of the staring experiment. In the present study, students of

psychology were very suspicious and many of them thought that finding the staring effect would be a mere cover story. For example, one of the students suggested that the "real aim" of the experiment was the observation of his posture. Thus, a description of the phenomenon has to be communicated very carefully.

Third, the methodology on staring studies must be expanded. For example, the length of test trials: Sheldrake (2000) said that 10 seconds were enough. However, Colwell *et al*. (2000) implemented 20 sec-trials, Lobach and Bierman (2003) implemented 30 sec-trials with 10 sec-inter-trial intervals. Most recently, Müller et al. (2006) implemented 1 minute-staring phases. However, instead of always responding at the end of a phase, participants could press a button whenever they felt stared at during the whole experiment. But how do we know what is long enough? If the staring effect is physiologically mediated, then how long does it last until a staring signal elicits a perceivable physiological reaction? Does the arousal immediately stop when the staring signal stops? What about alternative explanations? How can they be tested?

Fourth, instead of conducting a two-alternative forced choice task (i.e., a yes/no task), a confidence rating experiment would be sufficient to allow for different response biases and to capture the phenomenon in more detail. According to SDT, the task would be to respond to *each* trial how confident one is that someone has sent a staring signal (e.g., on a five point scale). If the remote staring effect exists, higher confidence ratings (and higher sensitivity scores at the end of the confidence scale) are expected on staring trials than on non-staring trials. Thus, participants were no longer forced to respond in favour of or against a staring signal, but had the opportunity to describe their sensations in a more detailed manner (see Macmillan & Creelman, 2005, for further details about confidence rating experiments).

Finally, most researchers in the field pursue an inductive research strategy, but in the long run, theoretical considerations will be needed that will help us to justify our methodology more accurately (e.g., a recently discussed theoretical approach is Generalized Quantum

Theory; for an overview, see von Stillfried, 2008; von Lucadou, Römer, & Walach, 2007).

References

Atkinson, A. P. (2005). Staring at the back of someone's head is no signal, and a sense of being stared at is no sense. *Journal of Consciousness Studies, 12*, 50-56.

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

Braud, W., Shafer, D., & Andrews, S. (1993). Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection. *Journal of Parapsychology, 57*, 373–90.

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.

Coover, J. E. (1913)."The feeling of being stared at"—experimental. *American Journal of Psychology*, *24*, 570–575.

Honorton, C. (1977). Psi and internal attention states. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 435-472). New York: Van Nostrand Reinhold.

Krampen, G. (2004). Differentielle indikation von autogenem training und progressiver relaxation. *Entspannungsverfahren*, *21*, 6-27.

Lobach, E., & Bierman, D. J. (2004). The invisible gaze: Three attempts to replicate Sheldrake's staring effects. *Proceedings of Parapsychology Association Annual Convention*, 77-89.

Lucadou, W. v., Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *Journal of Consciousness Studies, 14*, 50-74.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2.Ed)*. Cambridge: University Press.

Müller, S., Schmidt, S., & Walach, H. (2006). Remote staring detected by conscious and psychophysiological variables: Combining and improving two successful paradigms. *The Parapsychological Association Convention*, 85-99.

McNicol, D. (2005). *A primer of signal detection theory*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Poortman, J. J. (1959). The feeling of being stared at. *Journal of the Society for Psychical Research*, *40*, 4–12.

Radin, D. (2004). The feeling of being stared at: An analysis and replication. *Journal of the Society for Psychical Research*, *68*, 245–252.

Radin, D. (2005). The sense of being stared at: A preliminary meta-analysis. *Journal of Consciousness Studies, 12*, 95-100.

Schlitz, M., Wiseman, R., Watt, C., & Radin, D. (2006). Of two minds: Sceptic-proponent collaboration within parapsychology. *British Journal of Psychology*, *97*, 313-322.

Schmidt, S. (2005). Comments on Sheldrake's 'The Sense of Being Stared At'. *Journal of Consciousness Studies, 12*, 105-108.

Schmidt, S., Schneider, R., Utts, J., & Walach, H. (2004). Distant intentionality and the feeling of being stared at: Two meta-analyses. *British Journal of Psychology, 95*, 235-247.

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

Sheldrake, R. (2000). *The staring experiment*. http://www.sheldrake.org/Onlineexp/offline/staring_experiment.html.

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

Sheldrake, R. (2005). The sense of being stared at. Part 1: Is it real or illusory? *Journal of Consciousness Studies, 12*, 10-31.

Stillfried, N. v. (2008). A systems-theoretical generalization of non-local correlations. In S. Vrobel, O. E. Rössler, & T. Marks-Tarlow (eds.), *Temporal structures and observer perspectives*. World Scientific: Singapore.

Titchener, E. B. (1898). The feeling of being stared at. *Science New Series*, *8*, 895–897.

Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology*, *61*, 197–207.

Wiseman, R., & Schlitz, M. (1999). Replication of experimenter effect and the remote detection of staring. *Proceedings of the 42nd Annual Convention of the Parapsychological Association*, 471–479.

European Journal of Parapsychology
Volume 24.1, pages 53-67

ISSN: 0168-7263

A Controlled Long-Distance Test of a Professional Medium

Christian Gaden Jensen*† & Etzel Cardeña†

* Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Denmark
† Center for Research on Consciousness and Anomalous Psychology (CERCAP), Department of Psychology, Lund University, Sweden

Abstract

Suitable methods for testing alleged mediums are still debated after a century of research. In this study a professional medium was tested using a double-masked, long distance protocol with seven male sitters who rated how each statement and overall readings applied to them; they also completed a measure of paranormal belief. The experimenters rated the specificity of the statements. Statement specificity was negatively correlated with applicability, whereas paranormal belief was positively related to overall applicability ratings, but not to sitters' ratings of their target reading. No sitter rated his target reading as the most applicable and the statistical analysis based on the Pratt and Birge (1948) technique did not support the hypothesis of genuine mediumistic ability. Possible reasons for these results are discussed as are methodological issues in the quantitative assessment of mediumship.

Introduction

Methods for testing alleged mediums have been the centre of much discussion and controversy for a long time and are reviewed comprehensively elsewhere (Braude, 2003; Fontana, 2005; Gauld, 1984; Schouten, 1994). O'Keeffe and Wiseman (2005) argue that the

Correspondence details: Etzel Cardeña, Thorsen Professor of Psychology, Department of Psychology, Lund, University, P. O. Box 213 SE-221 00 Lund, Sweden. Email: Etzel.Cardena@psychology.lu.se.

methodological problems have centred around three key issues: (a) the need to control for sensory leakage, (b) the need to accurately assess the generality of the mediums' statements, and (c) the need for masked judgement. As a response to these shortcomings, they developed a controlled procedure for testing mediums. Controls for sensory leakage should prevent the medium from gaining information about sitters via normal means; the medium should not be able to ascertain information on the sitters prior to the reading, and verbal and non-verbal cues should be prevented during the reading. The assessment of the generality of the statements has been much debated, but historically probably the most widely employed procedure is that of Pratt and Birge (1948), in which a small number of sitters receive a reading from a medium and are asked to blindly rate both the reading intended for them (the *target* reading) as well all the other readings (the *decoy* readings). The full Pratt and Birge protocol is employed to a lesser extent today, but central elements of the procedure were employed here and are described further below. The need to ensure proper masked judging arises from observations that sitters rate mediumistic readings in very subjective ways (Hyman, 1977; Wiseman & O'Keeffe, 2001) because of personal beliefs and psychological needs, selective recall, forgetting, wishes to please the medium, and other reasons. This subjective rater bias is not totally eliminated by the Pratt and Birge procedure since subtle temporal cues (e.g., mentions of the weather, time of day, public events) in the statements might reveal at what day or time the reading was given, allowing the *recipient* (sitter for whom the reading is intended) to infer if s/he is the recipient or not (O'Keeffe & Wiseman, 2005). To overcome this difficulty, O'Keeffe and Wiseman introduced counterbalanced scheduling of the sitters' readings.

Robertson and Roy (2001; 2004; Roy & Robertson, 2001) have carried out various experiments with alleged mediums, varying the circumstances of the sessions. They have presented supportive evidence that the mediums were able to acquire ostensibly veridical information since recipients significantly accepted the mediums' statements as more relevant to their lives than did non-recipients. The

effects seemingly could not be explained as the result of verbal or non-verbal cues or sitters' knowledge of whether they were the recipients or not. The Robertson-Roy Protocol (Roy & Robertson, 2001) in full form definitely seems worth exploring further, but it is not designed for long distance experiments and it is time consuming.

Emily Kelly (2007) recently used proxy sittings to control for sensory leakage in tests of alleged mediums. In the second phase of this experiment, which yielded significant results, she or her colleague acted as proxy sitters and recruited the real sitters from among people they knew. Kelly states that the proxy sitters knew little about the deceased individuals and sitters. The procedure involved presenting mediums with the name, birth date, and a photograph of the deceased individuals. Sensory leakage through the verbal responses of the proxy sitters to readings of the few people they did know beforehand cannot be ruled out since one of the sitters was not masked to the photos, but a comparison between the unmasked and masked proxy sitters revealed that both obtained significant results. Although the chosen photographs were considered 'neutral' and did not involve specific activities, it is possible that the photos allowed for the use of cold reading techniques, and sending them to the mediums in advance reduces the control for possible fraud, however unlikely. Kelly also edited the transcripts of the sessions to remove fillers and identifying information derived from the photos before the sitters actually chose which transcript referred to them. Using a masked editor would be a methodological improvement over this procedure.

Gary Schwartz and colleagues (Schwartz *et al.* 2001; Schwartz, 2003) have also conducted a series of experiments with mediums and concluded that they obtained convincing evidence for the survival hypothesis. The experiments have been criticized (Hyman, 2003; 2003b; Wiseman & O'Keeffe, 2001) for not being double masked, for choosing sitters with a disposition towards the survival hypothesis, for lacking controls for sensory leakage from the recipients, and for allowing subjective rater biases to influence the mediums' accuracy scores. Schwartz replies (2001; 2003b) that the research team chose more naturalistic experimental designs to develop professional trust with the

mediums and to examine how mediumship is often conducted in the field. They did not intend to eliminate these possible explanations, but to minimize them in an exploratory phase, and slowly work towards more controlled experiments, in which they have continued to get significant results (Beischel & Schwartz, 2007). As discussed later, a more naturalistic approach may present certain advantages (Beischel, 2007).

This experiment was tightly controlled and adds some procedural refinements to the studies described previously. Sensory leakage was prevented by the use of a double-masked, long-distance design. Statement specificity was rated by the experimenters under masked conditions and sitters rated the extent to which all statements and readings applied to them. The medium's reported confidence in each reading was compared to the applicability ratings and transcriptions were assembled in sets counterbalanced as to both order and status (target or decoy) for each sitter.

To evaluate the possibility of genuine, mediumistic ability, this experiment tested four predictions. The first prediction was that target readings would receive greater applicability ratings than decoy ones. The second held that statement specificity would be negatively correlated with averaged sitter applicability ratings but positively correlated with the recipient's applicability rating, since highly specific statements should more often fit the recipient than any other sitter. The third prediction was that sitters' average ratings would correlate with paranormal belief and confirm the notion of a subjective rater bias. The fourth, exploratory, prediction was that the medium's self-reported confidence levels would be related to recipients' applicability ratings. This would corroborate the medium's own reports obtained through personal communication as well as other professional mediums' phenomenological reports of their ability to feel when there is a 'reliable' connection with a putative paranormal information source (e.g., Dampier-Jeans, 2001; Gater, 1995).

Method

Participants

The medium: The anonymous medium is internationally renowned and is regarded as an excellent medium in spiritualist circles. For over twenty years she has worked extensively as a medium in a variety of private and public settings. She provided informed consent and did not receive any compensation. She had not previously participated in any controlled experiments of her alleged mediumship ability.

The sitters: Seven men ranging in age from 28 to 68 years (M = 38.00, SD = 15.55) participated as sitters and provided informed consent. None reported having been to a medium before. All had lost at least two *significant* persons, defined as "someone with whom you had a personally important relationship, not just someone you knew superficially." Three of the seven had lost at least one *close* person, defined as "someone you loved, shared your innermost secrets with, or cared deeply about." Six sitters had lost a significant or close person within 10 years and four within 5 years.

Materials

The *Revised Paranormal Belief Scale* (RPBS; Lange, Irwin, & Houran, 2000; Tobacyk, 1988; Tobacyk & Milford, 1983) was used as a measure of paranormal belief. It has twenty-six items anchored on a seven-point Likert scale and two scales that measure traditional paranormal (TPB) and new age (NAP) beliefs.

For audio recordings we used a Mini Disk SHARP Digital Audio Recorder, MD-MT20. This unit performs by 24-bit ATRAC encoding and has an adjustable record-level, which can reduce low-frequency noise. An external microphone on a stand was connected to the MT20.

For video and secondary audio recording we used a Sony DCR-IP7BT digital camcorder that records MPEG-2 video. All recordings were of satisfactory quality.

Procedure

Experimenters 1 and 2 correspond to the authors of this paper, Devin Blair Terhune acted as experimenter 3. Throughout the experiment, E1 was masked as to the identity of the sitters, whereas E2 and E3 had no contact with the medium. No information concerning the medium or the time or the place of the readings was disclosed to the sitters. No information other than the name of each sitter was provided to the medium at the time of the respective reading. Finally, the sitters were chosen to ensure that none had unique names and the names were ethnically similar (Swedish-sounding). However, some information (e.g., sex, age) may have been inferable from the names, as discussed below. Sitter names were individually enclosed in sealed envelopes labelled with random numbers from one to seven by E3.

The readings were supervised by E1 in Karlstad, Sweden, at a local mediumistic school, chosen by the medium because she was giving a course there. E1 conducted a thorough visual search of the room and surrounding rooms for hidden cameras and microphones and covered up the windows. The medium herself was visually inspected, but not thoroughly searched. She sat in a chair behind a table with no tablecloth on which the microphone was placed on a stand. The video recorder was placed on a tripod at a distance of about three meters from the medium, ensuring the visibility of her full figure. Recording was initiated, the instructions were repeated, and she was given the large, sealed envelope. E1 left the room and closed the door.

After getting into the altered state of mind she uses during seances, and of which she gave verbal reports at a later time, the medium opened the large envelope and followed the same procedure for each reading. She opened a small envelope prepared by E3 and silently read the name of the sitter and wrote the name down behind a plastic screen, so the name could not be seen on the video recording. She began and ended each reading by stating the respective reading number. She was instructed to not state the name of the sitter aloud, and she adhered to this instruction. Following the completion of each reading, she provided a self-report of her confidence in the preceding

reading (1: *Unconfident*; 2: *Somewhat unconfident*; 3: *Somewhat confident*; 4: *Confident*). The allotted time for each reading was 15 minutes to avoid tiring the medium, since she was asked to give 7 readings in one day. Each reading was timed by E1 and the medium was notified of the time remaining at 12 and 14 minutes into each reading. The medium was encouraged to take short breaks between each reading to avoid overflow of information from the preceding reading, and to "follow her natural rhythm," rather than sticking to a controlled schedule.

E1 subsequently examined the video and found no behaviour suggestive of chicanery or fraud. He transcribed the readings, removed extraneous details (e.g., comments at the beginning and end of a reading), and segmented them into series of statements (SOS). This procedure adhered to that used by O'Keeffe and Wiseman (2005). The full text and SOS for each reading were blindly examined and corroborated by E2 and E3. Each statement was rated for its specificity (1: *Overgeneral*; 2: *Somewhat general*; 3: *Somewhat specific*; 4: *Very specific*) by the experimenters, who did so masked to each other's ratings.

E1 next designed one rating set specifically for each sitter, with all seven SOS in each, placed in a double-counterbalanced order, which took into account both the placement of the target reading and the order of all the readings and was done to ensure that each sitter received a rating set with his target reading at a different position. Each sitter first read all seven readings in full text to give an overall, meaningful understanding of each reading; these were followed by the seven sets of SOS.

The sitters completed the RPBS and provided ratings for individual statements and global ratings for each SOS. Participants were instructed to rate each individual statement for its applicability to their personal history and current life situation (1 [*Not applicable*] to 7 [*Very applicable*]) and globally rate the applicability of each SOS (1 [*Not applicable*] to 100 [*Very applicable*]). Participants were further instructed to briefly report the reasons why they gave their highest rating to a particular reading. Data were computed by E3 while he remained masked to the target readings.

Statistical analyses

To test whether the recipients' ratings were higher than expected by chance, a permutation analysis similar to the one employed in O'Keeffe and Wiseman (2005) was carried out. The Pratt and Birge technique (1948; Pratt, 1969) from which the analysis originates was designed to handle free-response data from psychic readings. All responses (in this case ratings) are arranged in a matrix with responses for targets placed diagonally through the matrix, and then whether the sum or the average of the numbers in the diagonal cells deviate significantly from a chance distribution is determined. When the sample size is below 10 (as it is here), it is possible to calculate this probability permutation analysis in which a computer program rearranges the numbers in the matrix in every possible way, computing the score (the sum or the average) of the diagonal cells every time. Statistical significance occurs if the proportion of the scores that are more extreme than the score in question is less than the criterion *p*-value (Edge *et al.*, 1986; for a detailed discussion of this analysis, see also Greville [1944] on which Pratt and Birge based their work, see also Greville, 1949; Scott, 1972; and Thouless, 1949).

Results

The readings ranged in duration from 5:01 to 12:45 min. (M = 8:41, SD = 2:55) and ranged in number of statements from 14 to 61 (M = 30.3, SD = 14.9). The medium did not report fatigue associated with the procedure and provided the relatively highest level of confidence in, and greatest number of statements for, the last reading.

Applicability Ratings

In contrast with the first prediction, no sitter chose his target reading as the most applicable. Target readings were ranked 3rd, 3rd, 3rd,

5th, 6th, 7th, and 7th as measured by the recipients' overall scores[1] (see Table 1). They were ranked 2nd, 2nd, 3rd, 5th, 6th, and 7th as measured by the recipients' average ratings of the SOS (see Table 2). Target rank derived from final rating and target rank derived from average statement ratings correlated significantly, indicating that the sitters' overall evaluation for each reading corresponded to their average rating of the single statements in each SOS (*rho* = .89, *p* = .008). Performing 10,000 permutations in a computer programme for the overall scores of the recipients yielded a *z*-score of -1.2, *p* = .89. Ten thousand permutations for the recipients' average statement ratings resulted in *z* = -1.3, *p* = .90. Thus, neither the overall scores nor the average statement ratings for the target SOS were significantly better than expected by chance. These results are not supportive of anomalous cognition during the experiment.

Confidence ratings

The medium's confidence ratings were relatively low with five ratings of 1 (unconfident) and two of 2 (somewhat unconfident). Contrary to the exploratory hypothesis, the medium's confidence ratings correlated negatively with the recipients' ratings of the readings: *rho* = -.80, *p* = .032, and non-significantly with the sitters' average statement ratings: *rho* = - .16, *p* = .74. The former correlation may be suggestive of psi missing; however, considering the small sample size and the low variance in the medium's confidence ratings, this should not be considered a crucial finding. Also, the medium's confidence levels did not correlate with the number of statements in a reading: *rho* = .32, *p* = .49.

[1] When overall scores (contrary to instructions) were the same for two readings in a set (*n* = 2) the higher overall rank between the two was assigned to the one with the higher average statement score.

Table 1. Sitters' overall applicability ratings (target ratings in bold)

	Reading							
Sitter	A	B	C	D	E	F	G	Total
A	**22**	25	10	15	20	23	7	122
B	15	**17**	15	50	12	20	10	139
C	32	82	**19**	29	47	61	31	261
D	15	75	10	**20**	30	50	45	245
E	50	55	70	60	**56**	45	40	376
F	15	32	27	25	20	**19**	30	621
G	6	10	4	5	11	9	**4**	49
Total	155	296	155	204	196	227	167	157

Note. The value of 157 represents the sum of the ratings along the diagonal.

Table 2: Sitters' average statement ratings (target ratings in bold)

	Reading							
Sitter	A	B	C	D	E	F	G	Total
A	**2.55**	2.74	2.21	1.94	2.5	2.48	2.03	16.45
B	3.36	**3.22**	2.41	4.09	2.93	2.9	2.28	21.19
C	3.59	5.83	**2.79**	3.78	3.57	4.35	4.07	37.64
D	3.09	4.87	2.14	**2.81**	3.71	3.58	3.82	24.02
E	4.64	4.48	5.34	4.56	**4.64**	4.58	3.93	32.17
F	2.91	4.43	3.34	3.56	2.43	**2.97**	3.84	56.19
G	1.77	2.22	1.55	1.72	2.14	1.77	**1.43**	12.6
Total	21.91	27.79	19.78	22.46	21.92	22.63	21.4	20.41

Note. The value of 20.41 represents the sum of the ratings along the diagonal.

Paranormal belief

The sitter's scores on the TPB (M = 25.15, SD = 7.15) and NAP (M = 23.94, SD = 3.16) measures were within the normal range for non-clinical samples (see Lange et al., 2000). The two measures were significantly correlated, *rho* = .79, p = .036. Average ratings significantly correlated with TBP (*rho* = .89, p = .007) and NAP (*rho* = .78, p = .04). This indicates that paranormal believers judged statements to fit their lives to a higher degree than non-believers. Neither subscale correlated with the target SOS overall ranking (TPB: *rho* = .06, p = .91; NAP: *rho*

=.19, p = .69), or rank derived from average statement ratings (TPB: *rho* = -.29, p = .53; NAP: *rho* = .07, p = .88.).

Statement specificity

The three experimenters' inter-rater reliability concerning the specificity of the statements was adequate (Spearman *rho*s ranged from .46 to .53, $p < .001$). A significant, negative correlation was found between statement specificity and average rating from all sitters combined: *rho* = -.38, $p < .001$. However, in contrast with our prediction, the more specific a statement was, the *lower* the rating it received from the recipient (n = 212; *rho* = -.12, p = .073).

Discussion

Methodologically this experiment was strictly controlled. The experimental conditions permitted a very small possibility of sensory leakage regarding sitters' names, the distance between the medium and the sitters was 800 kms, the principal experimenter (E1) was masked to the sitters' identities, and the secondary experimenters (E2 and E3) had no interaction with the medium and were masked to the target readings during sitter rating sessions. The sitters received no information about the medium, who was only presented the names of the sitters, the rooms were carefully inspected to minimize the potential for fraud, and the whole experiment was recorded on both video and audio. Some information may have been inferable from the sitters' names, but this was unlikely because the medium was dealing with persons of a foreign nationality where naming trends were probably unknown to her. However, ruling out this confound seems impossible within this methodological framework.

The generality of the medium's statements were accurately assessed: each sitter rated all given statements and all given readings as a whole and all three experimenters rated all statements for their specificity. The sitters also rated the SOS masked as to which was their

target reading and the SOS were administered in double-counterbalanced order.

Considering psychological variables, the medium seemed highly motivated to obtain recognizable information, her belief in mediumship is very high, and she reported being in her usual altered state of consciousness. These factors could normally be expected to contribute to a heightened chance for anomalous information transfer, as they often correlate with more positive scores on ESP tests (Palmer, 1997). On the other hand, all readings were given low confidence ratings by the medium; thus, it is apparent that this protocol did not provide an environment conducive to the medium's confidence in her ability to obtain what she usually feels to be a "reliable" connection with a paranormal information source. Based on the research of the Windbridge Institute for Applied Research in Human Potential, Beischel (2007) argues that mediumship research should emphasize not only the experimental masking of the medium, the experimenters and the raters, but also aim for a research environment that optimizes the mediumship process for both the medium and the hypothesized discarnate. Two possible improvements of the highly controlled method applied here therefore consist in supplying the medium with the name of one specific discarnate for each sitter, rather than the name of each sitter and in screening the included discarnates to ensure different personalities and causes of death, These changes more closely mimic what is often the format of a "natural" reading, and may help focus the medium. Furthermore, as Fontana (2005) points out, trial runs with mediums, of which this experiment had none, may be crucial to develop optimal experimental methodologies and find the mediums who appear able to perform well under the chosen conditions.

The finding that paranormal believers judge statements to fit their lives to a higher degree than non-believers support the notion that judging the accuracy of a reading is a highly subjective matter (Hyman, 1977; Wiseman & O'Keeffe, 2001), and replicate earlier findings (O'Keeffe & Wiseman, 2005; Robertson & Roy, 2004).

One potential limitation of this study is contextual; each statement was always presented to participants in the context of the series of

statements proffered by the medium in the respective reading. This may facilitate a sitter bias in which two statements of relatively equivalent applicability drawn from different readings receive differential ratings due to the applicability of the surrounding statements. Foregoing the administration of statements in reading-based clusters and randomizing them may improve upon the internal validity of the experiment by reducing the potential for this contextual sitter bias. However, this sacrifices ecological validity since the meaningfulness (i.e., applicability) of a statement often resides precisely in the context. Thus, when measured quantitatively, the whole may well constitute more than the sum of its parts as judged in isolation. Future studies should consider this issue, which has been a recognized methodological challenge as early as Saltmarsh & Soal (1930) and in Hyslop's (1919) work on probability of statements.

This study is naturally too small to draw any firm conclusions, but we think that the adopted protocol circumvented the methodological shortcomings that have plagued previous studies of mediums (see O'Keeffe & Wiseman, 2005). Improvements may lie in administering controls for sensory leakage in a way that does not allow for normal information transfer but at the same time allows for entanglement or 'organizational closure' to occur, as suggested by von Lucadou, Römer, and Wallach (2007). This is highly speculative, but the medium in this experiment specifically mentioned the impossibility of "blending into" the mind of the sitter as a reason for her low confidence levels. Establishing a person-to-person contact at long-distance without losing the control for sensory leakage might be a worthwhile methodological challenge to accept for long distance studies of mediumship.

Acknowledgements

We gratefully acknowledge the participation of Devin Blair Terhune, who acted as an assistant experimenter, an initial critical reading by Nils Wiklund, Ph.D., of our preliminary design, and Dean Radin, Ph.D., for creating the computer program carrying out the numerical simulations in the Pratt-Birge procedure.

References

Beischel, J., & Schwartz, G. E. (2007). Anomalous information reception by research mediums demonstrated using a novel triple-blind protocol. *Explore, 3,* 23-27.

Braude, S. E. (2003). Immortal remains: The evidence for life after death. Lanham, MD: Rowman & Littlefield.

Dampier-Jeans, M. (2001). *Mit liv med ånderne*. København; Borgens Forlag.

Edge, H. L., Morris, R. L., Palmer, J., & Rush, J. H. (1986). *Foundations of parapsychology: Exploring the boundaries of human capability.* London: Routledge and Kegan Paul.

Fontana, D. (2005). Is there an afterlife? A comprehensive review of the evidence. Oakland, CA: O Books.

Gater, D. (1995). *A psychic's casebook*. London: Robert Hale Ltd.

Gauld, A. (1984). Mediumship and survival: A century of investigations. Chicago: Academy Chicago Publishers.

Greville, T. N. E. (1944). On multiple matching with one variable deck. *Annals of Mathematical Statistics, 15,* 432-434.

Greville, T. N. E. (1949). On the number of sets required for testing the significance of verbal material. *Journal of Parapsychology, 13*, 137-138.

Hyman, R. (1977). Cold reading: How to convince strangers that you know all about them. *Skeptical Inquirer, 1*, 18-37.

Hyman, R. (2003). How not to test mediums: Critiquing the afterlife experiments. *Skeptical Inquirer, 27*, 20-30.

Hyman, R. (2003). Hyman's reply to Schwartz's 'How not to review mediumship research'. *Skeptical Inquirer, 27,* 61-64.

Hyslop (1919). *Contact with the other world*. NY: Columbia University Press.

Kelly, E. W. (2007). *Investigation of mediums who claim to give information about deceased persons*. Unpublished Progress Report prepared for The Bial Foundation.

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.

O'Keeffe, C., & Wiseman, R. (2005). Testing alleged mediumship: Methods and results. *British Journal of Psychology, 96*, 165-179.

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). NY: Parapsychology Foundation.

Robertson, T. J., & Roy, A. E. (2004). Results of the application of the Robertson-Roy Protocol to a series of experiments with mediums and participants. *JSPR, 68*, 18-34.

Robertson, T. J., & Roy, A. E. (2001). A preliminary study of the acceptance by non-recipients of mediums' statements to recipients. *JSPR, 65*, 91-106.

Roy, A. E., & Robertson, T. J. (2001). A double-blind procedure for assessing the relevance of a medium's statements to a recipient. *JSPR, 65*, 161-174.

Saltmarsh, H. F., and S. G. Soal (1930). A method of estimating the supernormal content of mediumistic communications. *Proceedings of the Society for Psychical Research*, *39*, 266-271.

Schwartz, G. E. (2003). *The afterlife experiments: Breakthrough scientific evidence of life after death.* NY: Atria Books.

Schwartz, G. E. (2003b). How not to review mediumship research. *Skeptical Inquirer*, *27*, 58-61.

Schwartz, G.E. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A call for balanced evidence-based skepticism. *The Paranormal Review: 20*.

Schwartz, G.E., 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*: 1-25.

Scott, C. (1972): On the evaluation of verbal material in parapsychology: A discussion of Dr Pratt's monograph. *JSPR, 46*, 79-90.

Schouten, S. A. (1994). An overview of quantitatively evaluated studies with mediums and psychics. *Journal of the American Society for Psychical Research, 88*, 221-254.

Thouless, R. H. (1949): Review of Journal of Parapsychology, Vol. 12, Number 4. *Journal of the Society for Psychical Research, 35*, 48-49.

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.

von Lucadou, W. Römer, H., & Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *Journal of Consciousness Studies, 14*, 50–74.

Wiseman, R., & O'Keeffe, C. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A critique. *Paranormal Review*, *19*, 3-6.

European Journal of Parapsychology
Volume 24.1, pages 68-92

ISSN: 0168-7263

Personality, Mental State, and Procedure in the Experimental Replication of ESP: A Logistic Regression Analysis of a Successful Experimental Condition

José M. Pérez-Navarro*, Tony Lawrence**
and Ian Hume**

* Department of Psychology and Counselling, University of Greenwich, United Kingdom
** School of Health and Social Sciences, University of Coventry, United Kingdom

Abstract

Despite many years of research, evidence in support of ESP (extrasensory perception) remains controversial, mainly, due to the difficulty in replicating positive results across studies. This study explores a series of variables as predictors of study success. Ninety participants replicated an experimental condition, based on the use of Ganzfeld and multi-sensory stimuli, which had produced significant results in a previous study. These participants achieved a significant hit rate of 33.3% (z=1.86, p=.031, one tailed). Variables psychoticism, pre-session tense arousal, anger and frustration, empathy, feeling upset, discomfort, and concern about the external environment during the Ganzfeld correlated positively with the participants' performance (p<.05). Conversely, variables internal awareness, vulnerability, feeling vigorous, and sender-receiver intimate bonds showed negative associations. A logistic regression analysis of these predictors showed a four-variable equation that predicted correctly 79.55% of the cases. We concluded that the use of Ganzfeld and multi-sensory targets in the experimental testing of ESP is worth further exploration.

Correspondence details: Dr José M. Pérez-Navarro, Department of Psychology and Counselling, University of Greenwich, Avery Hill Campus, Eltham, London, United Kingdom, SE9 2UG. Email: j.m.pereznavarro@gre.ac.uk.

Introduction

Researchers in the area of parapsychology have made great efforts to replicate in the laboratory a series of anomalous phenomena and experiences observed in spontaneous case reports. In the case of ESP (extrasensory perception), in particular, since the first experimental work conducted by J. B. Rhine at Duke University in the 1920s, a large number of studies have been conducted. The experimental procedure most commonly used in recent times to study ESP is the Ganzfeld technique (see Bem & Honorton, 1994). Although, results from studies using this technique vary in their degree of support for ESP, meta-analyses show a small but highly significant effect of information transfer between a sender and a receiver that supports the existence of a process of ESP communication (Bem & Honorton, 1994; Storm & Ertel, 2001). However, some researchers criticise the lack of replication of positive results consistently across studies (e.g., Milton and Wiseman, 1999, 2001, 2002). Bem, Palmer, and Broughton, R. (2001) argue that this heterogeneity of outcomes could be attributed to researchers not conforming to the original Ganzfeld protocol. Nevertheless, the lack of understanding of the underlying mechanisms of this phenomenon and the factors its occurrence and scale rely on has not allowed researchers to develop a robust experimental protocol to replicate the ESP phenomenon consistently and visibly across laboratories (Perez-Navarro, 2003).

Along this line, researchers have explored a large number of personality traits, mood, situational, and interpersonal variables in an effort to outline a 'recipe' for successful ESP testing. Traits like *extraversion* or *prior laboratory testing* may relate merely to the ability to be at ease in the testing situation, while an interest in or *practice of mental disciplines* may reflect a general interest in inner experiences. Other factors such as *personal psi experiences* and high scores on the *feeling and perception* poles of the Myers-Briggs Type Indicator may be more central in the ESP process (Parker, Grams, & Petterson, 1998). Many of the variables explored can be understood in relation to the Noise Reduction Model (NRM, Honorton, 1977, 1978). In the NRM,

ESP is conceptualised as a weak signal that is frequently masked by internal somatic and external sensory 'noise'. Reducing the noise-to-signal ratio should therefore help detect any psi signal, and this can be achieved by reducing internal and external stimulation. According to this, *relaxation* is one of the most desirable condition since it could be a means of enhancing the signal-to-noise ratio by reducing somatic and cognitive noise in ESP experiments. The experimental evidence, however, is not as clear as can be expected from the theory. Several researchers report a positive association between the participants' performance and their degree of relaxation, (see for example Braud & Braud, 1973; Sargent, 1982; or Stanford & Mayer, 1974). However, other authors have failed to find a significant association between these two variables (e.g., George, 1982; Morris & Morrell, 1985; Musso & Granero, 1982). Braud (1977) found a curvilinear relationship between relaxation and ESP scores suggesting that an optimum level of arousal could be required in this type of experiment. If this was the case, traits like *neuroticism* or *anxiety*, for example, would help the participant to maintain arousal if tested in a too dormant environment. Conversely, the same traits could turn counterproductive interfering the process of relaxation in situations somewhat more stimulating. Thus, indicators of the individual's physical and mental state during the session such as *mood*, or simply *comfort* experienced could help us to predict the session outcome.

The NRM suggests that techniques used to encourage participants to direct their attention to internal sources of information could help in the detection of the ESP signal. However, individuals might differ naturally in their tendency to do this. For example, *internal awareness* is defined as the degree of awareness of one's own thoughts, motives, and other internal aspects of the self (Fenigstein, Scheines, & Bush, 1975). Thus, it could be hypothesised that individuals high on this trait would notice any ESP information or ESP-mediated imagery easier. Similarly, once the ESP information is detected, the NRM suggests that this may appear in the individual's consciousness as a weak signal masked by cognitive, physiological and physical noise. Therefore, the individual might experience this information as unclear and

unstructured. In this regards, Kruglanski, (1990, p. 337) defines the construct *need for cognitive closure* as a desire for "an answer on a given topic, any answer, compared to confusion or ambiguity". Therefore, participants with a high need for closure could cope with the ambiguity of this information by processing it on the basis of their own mental schemas. This, in principle, could be detrimental to the task as would distort further an already weak signal. This is also related to Gordon's (1949) concept of *controllability of mental imagery*. Gordon refers to individuals as controllers, if they are able to control and manipulate their mental imagery, or autonomous, if they are not. Such manipulation, if given during the experiment, could distort weak ESP-mediated imagery.

Other variables appear in the literature associated to the participants' performance. For example, Haraldson (1985) has explored extensively the construct *perceptual defensiveness* in relation to the ESP task, finding that those participants who score higher in the Defence Mechanisms Test (Kragh, 1960, 1986) also tend to succeed in forced-choice ESP tests. The work of other authors on this construct and psi (Watt, 1992; Watt, & Ravenscroft, 1999), however, is less encouraging. Similarly, other researchers have explored the construct *creativity*. However the diversity of measures used in these studies leaves us no conclusive answer to whether this ability is related to ESP. Some studies (e.g., Dalton, 1997), though, report encouraging results.

Paranormal belief is one of the variables that have appeared most consistently related to the participants scores in ESP tests. A meta-analysis by Lawrence (1993) suggests that believers in the paranormal tend to score higher in forced-choice ESP test that non-believers. This could be more concretely related to the participant's perception of the ESP task and their estimation of the probability of success. Some studies have explored the participants' expectancies of success in order to find out whether these would predict performance. For example, Taddonio (1974) separated her participants in two groups and manipulated their perception of the probability of success in an ESP experiment. This author reports higher scores by those participants whose expectations had been positively biased.

Naturally occurring ESP experiences are frequently reported between family members or close friends. Some laboratory work suggests that emotional bonds between sender and receiver might facilitate ESP (Bierman, Bosga, Gerding, and Wezelman, 1993; Broughton, Kanthamani, and Khilji, 1989). In relation to this we can refer to the concept of empathy. Empathy is a psychological process through which the individual gains understanding of others' emotions, feelings and points of view. Empathy is viewed as a trait in some models of personality (Eysenck and Eysenck, 1991) and individuals might differ in their tendency to engage in this process. We think that an ESP mechanisms could be involved in this process. Thus, we expect that individuals who report this experiences in personality measures also score high in the ESP task

Mainstream psychology has considered the effect of biological rhythms on behaviours as well as cognitive activity, mood states, and other psychological parameters. The menstrual cycle in females, one of the most powerful rhythms, has proved to affect a large variety of psychological and physiological processes. Schmitt and Stanford (1972), in a Ganzfeld study with 20 females, observed a nearly significant effect of menstrual cycle on ESP. Fifteen females were in the pre-ovulatory phase and achieved 11 hits. Five were in the post-ovulatory phase producing only 1 hit. However, the effect of menstrual cycle on ESP suggested in this study could be mediated through other psychological and state variables. As Schmitt and Stanford argue, this variable warrants further research and could show great usefulness in participant selection for future studies.

Characteristics of the target also seem to play a role in the outcome of ESP experiments. In the PRL series, for example, significantly higher scores were reported for trials where video clips were used instead of art prints (Bem and Honorton, 1994). Other studies that have used multi-sensory targets also reported successful results. For example, in one of the Maimonides dream studies (Krippner, Ullman, & Honorton, 1971) the researchers used an experimental design in which a multi-sensory experience was designed around the target after the participant had awakened. Although no control condition was used in

this study, these authors report highly successful results. This is comparable to remote viewing studies, where the agent is taken to a randomly selected site and experiences his surroundings. Some analyses suggest that remote viewing studies are successful (e. g. Hansen, Schlitz, & Tart, 1984; Milton, 1998), although not much work has been conducted so far in order to contrast directly the degree of success of this procedure with the Ganzfeld.

In a previous study (Perez-Navarro, Lawrence, and Hume, 2009) we explored a series of traits and mental state factors that, from a theoretical point of view, could play a role in the ESP process. We used two experimental conditions: Ganzfeld vs. sensory attenuation, and two types of targets: pictures vs. objects. However, the variables that best predicted our participants scores showed very little consistency across the different experimental conditions and target types. Although the overall rate of correct guesses (26%) did not differ significantly from the 25% expected by chance, we observed, as hypothesised, that those sessions using Ganzfeld and multi-sensory targets produced a significant 43% rate of right guesses. We concluded that, as suggested by Bem *et al*., (2001), differences in the experimental environments and procedures across studies could be the reason for the general lack of replication in the area and that the convergence of researchers into a standardised protocol was needed. We suggested that future research explores other potential predictors of participants' performance using Ganzfeld and multi-sensory targets.

In the present paper, we report a new study where we replicated this experimental condition with 90 volunteers. We explored participants' traits and mental state factors, through logistic regression analysis, in an effort to outline a set of variables able to predict the experimental outcome. We included variables from the literature that had appeared associated to the participants' performance in previous studies as described above. The hypotheses in this study were stated in the same direction of the associations found in the literature.

We also included in this study those variables that showed a p-value below 0.05 in the correlation analyses in our previous study. The relationship between these variables and the individuals' performance

were hypothesised in the same direction. That is, a positive association was hypothesised between the individuals´ performance and variables internal awareness, empathy, impulsivity, controllability of visual imagery, feeling vigorous prior to the experiment, and post-session confidence of success. Conversely, negative associations were hypothesised for variables need for cognitive closure and feeling upset during the experiment (Perez-Navarro, Lawrence, and Hume, 2009).

Psychoticism is sometimes defined in contrast to socialisation and related to non-conformity, aggression, inconsideration of social rules and convention, and impulsiveness (Eysenck, 1967). In Costa and McCrae (1992) Five Factor Model of personality this construct is split in three factors: (low) conscientiousness, (low) agreeableness, and (high) openness, this latter factor includes the traits imagination, artistic interest, emotionality, adventurousness, intellect and liberalism. It is possible that individuals high on psychoticism find it easier to undertake a non-conventional task like the Ganzfeld. This variable has appeared associated with higher scoring in previous research (Haraldson and Houtkooper, 1991).

Other variables like comfort experienced by the participants during the session, altered state of consciousness, or absorption (Palmer, Khamashta, and Israelson, 1979; Stanford, 1979; Stanford and Angelini, 1984; Tart, 2000) were also assessed as indicators of mental and physical relaxation. We also thought it would be interesting to assess the individual's concern about the external environment during the Ganzfeld stimulation as this could be a measure of the individual's attempt to focus on the source of information. We also explored the individual's feelings of anger and frustration during the session, from a more psychodynamic perspective. Participants usually come to the psychology laboratory with certain expectancies about the test itself and about the treat they are going to receive from the team of experimenters. Given the complexity of the Ganzfeld protocol and the nature of human social interaction itself, it is possible that some participants feel frustrated at some point during the session. In consonance with classic theories of frustration-aggression (Dollard *et*

al., 1939), it could be the case that these participants under-perform, or even psi-miss, as a form of covered aggression to the experimenter.

On the basis of previous research (e.g., Bem and Honorton, 1994) and our own previous study, it was hypothesised that the overall hit rate of the study would be significantly above chance (α = .05). All hypotheses referred to correlations among the variables, if not otherwise indicated, are stated at a significant level of .01 in this study. Any tighter adjustment of α would have been, from our view, too conservative increasing markedly the probability of type-II error (see O'Keefe, 2003).

Method

Design

Only one experimental technique (the Ganzfeld procedure) and one target type (objects) were used in this study. The relationship between the individual differences and mental state, variables and participants' scores in the ESP test were explored through correlation and logistic regression analysis. The sessions outcome (dependent variable) were defined using a nominal scoring method for each participant. The individual simply picked on a 'blind' basis the one of the four pairs of objects that most closely resembled his/her experience during the period of Ganzfeld stimulation. If the stimulus chosen was the one that the sender participant was trying to communicate one *hit* was counted. Otherwise, the trial was coded as a *miss*.

Participants

Ninety participants were recruited through advertisement of the study amongst the student population at the university campus. These participants were enrolled in a variety of courses, though most of them were psychology students. Each volunteer was paid £5 in order to cover travelling and other expenses derived from their co-operation in the study. Individuals were encouraged to come along with a friend or

relative so that one could act as receiver and the other as sender. There were 36 males and 54 females. The age of the participants ranged from 18 to 43, with a mean of 24 and a standard deviation of 3.1.

Measures, apparatus, and materials

We used a clip-on microphone connected to a manual tape recorder to register the participants' report of their mental imagery and subjective impressions during the Ganzfeld sessions. A thirty-minute white noise track was created with the software CoolEdit and burned on a CD. White noise was generated to the receiver participant via headphones through a personal CD player. A wireless radio transmitter system set at the receiver's room fed back the receiver's mentation to the sender participant. The system received the input through the tape recorder and transmitted it to the sender's headset. Also a random number generator (RNG) was used to select the target stimulus among the pool of objects, as well as to randomise the order of presentation of the series of target and control stimuli to the receiver for judging after the session.

Target Stimuli: In this study we used a pool of sixteen pairs of objects organized into four sets (2 objects × 4 pairs x 4 sets). These were selected from a larger pool by the experimenter so that they could be interesting and attention-catching to the participants. They consisted, above all, of small toys, but also, small souvenirs, and daily utensils like key rings, a coffee cup, a biro, a CD-Rom, a bulb, a chocolate bar, a piece of soap, a hat, a pair of glasses, a small ball, etc. Each pair of objects was placed into a plastic bag, and each set (of four bags) was kept in a small box. Bags were labelled with the set number (a number from 1 to 4) and a letter from a to d for later random selection. The four boxes were labelled each with the set number they contained.

Questionnaires: The questionnaires listed next were used in the assessment of the individual differences and state variables. The development of the instruments are described in the referred

publications. The psychometric properties of the instruments appeared adequate in these studies. The instruments used were the following: Fenigstein's *Self-Consciousness Scale* (Fenigstein, Scheiner and Bush, 1975) for the assessment of internal awareness, Neuberg's *Need for Cognitive Closure Scale* (Neuberg, Judice and West, 1997), Riley's *Dissociation Scale* (Riley, 1988), Eysenck & Eysenck's *Impulsivity, Venturesomeness, Empathy Personality Inventory* (Eysenck and Eysenck, 1991), Gordon's *Controllability of Visual Imagery Questionnaire* (Gordon, 1949), *NEO-PI* (Costa and McCrae, 1992), *Keirsey Temperament Sorter* (Kersey, 1978), *The 18-Items Australian Sheep-Goat Scale* (Thalbourne, Delin, 1993), *Herrmann Brain Dominance Instrument* (Herrmann, 1999), Tellegen's *Absorption Scale* (Tellegen, and Atkinson, 1974), *Intimate Bonds Scale* (Wilhelm and Parker, 1988), *The UWIST Mood Adjective Checklist* (Matthews, Jones, and Chamberlain, 1990) and the NEO-PI questionnaire (Costa and McCrae, 1992) for the assessment of traits *neuroticism*, *extraversion* and *psychoticism*.

Perceptual defensiveness: The Defence Mechanisms Test (DMT, Kragh, 1960, 1985) has shown successful in the prediction of ESP scores in previous research (see e.g., Haraldsson and Houtkooper, 1995, for a meta-analysis). However due that this test is time consuming and requires prior intensive training, instead, a stroop task, based on the same reasoning as the DMT (i. e., a delayed or distorted perception of potentially threatening stimuli), was used in this study. Two sheets made of five columns of ten words each printed randomly on different colours were used. The first matrix was made of neutral words while the second was of emotionally threatening words. Participants were asked to name the colour of the ink of the words in the two matrices. Those participants who took longer with the threatening words were classified as defensive, while those who took less were classified as vigilant.

Creativity: An *alternate uses creativity task* (based on Wallach and Kogan, 1965) was designed in order to assess the participants' creativity. The individuals were given three objects, one after another,

in this order: a piece of cotton, a 50 cm ruler, and a party cone-shaped hat. They were asked to suggest possible uses of each of the objects and given one timed minute per object. The total number of uses given by the individual in one minute divided by three was taken as an individual's creativity indicator.

Other variables: A minor number of variables, like, *expectancies of success*, *prior testing*, *practice of mental disciplines*, *feeling vigorous* or *intake of stimulant drinks* prior to the session were assessed through five-point Likert-type items constructed by the experimenter. Similarly a post-session questionnaire also constructed by the experimenter was used in order to assess the individuals' feelings of *anger* and *frustration*, *discomfort*, or *concern about the external environment* during the sensory monotonization. The individual's *altered state of consciousness* experienced during the Ganzfeld stimulation was assessed through a series of indicators, like *feeling confused* or experiencing *time contraction* during the period of Ganzfeld stimulation, in a post-session questionnaire. *Phase of the menstrual cycle* was coded in females participants as 0 = pre-menstrual and 1 = post-menstrual.

Procedure

When prospective participants approached the experimenter (the author of this article) expressing their interest in the study, they were scheduled for two sessions a psychometric session and a Ganzfeld ESP test, on two different days. The psychometric session consisted of the administration of the questionnaires and assessment of the individual differences variables. The second session was a standard Ganzfeld ESP test.

Two experimenters were involved in the study: the first author of this article (experimenter A) and a co-experimenter (experimenter B). Experimenter A tested participants in the psychometric session and ran the ESP sessions. At the time of the session, experimenter A accompanied the receiver participant to the laboratory, on the 4th floor of the building, and asked him to fill in a pre-session questionnaire.

Meanwhile, experimenter B gave the instructions to the sender at the sending room, two floors above. Experimenter B, then, opened an envelope containing the code (randomly generated in advance) for target objects selection, giving the correspondent stimuli to the sender participant. Both experimenters were kept blind to these data until the time experimenter A left the room. At this time, experimenter A, in the laboratory, gave the instructions to the receiver in a standard manner, set up the radio transmitter and recording apparatus, and started the session.

Experimenter A remained all the time outside the receiver's cubicle, within the laboratory, listening to the individual's mentation through headphones and writing down his/her comments. In 30 minutes from the commencement of the session experimenter B, without knocking or producing any noise that could be regarded as a signal, passed a note under the laboratory door containing the set number of objects and a randomised sequence for presentation of the target and decoys. Experimenter A ignored this note until the time of the judging. After the 30 minutes of Ganzfeld stimulation experimenter A released the subject and proceeded to review his/her mentation. At this point the individual clarified and extended his/her mentation, adding any further details about his/her experience that considered important. Then, the participant was asked to fill in a post-session questionnaire about his subjective experience and mental state during the session. Next, experimenter A displayed on a table a duplicate of the set of objects that, according to the note passed under the door by experimenter B, contained the target objects the sender was attempting to communicate to the receiver. The individual was then asked to examine these four choices of objects, named A, B, C and D, and try to indicate which one resembled most closely his subjective experience during the period of sensory monotonization. At this time experimenter A was only aware of the set of stimuli that contained the target objects, but kept blind to which of these choices was the right one. It was a requirement of the protocol, at this point, that the experimenter would not help the individual in his decision in any way. Nobody at all was allowed to enter the laboratory until the

participant's response had been registered. Finally, when the judging process had been completed the experimenter accompanied the participant to the sending room to find out the identity of the target.

Results and Discussion

Hit rates and randomicity checks

From the 90 participants who took part in the study, 30 ranked the correct pair of objects in the first position, that is, as the most similar to the mental imagery and subjective impressions they experienced during the 30 minutes of Ganzfeld stimulation, producing a rate of successful guesses of 33.3%. The difference between this rate and the 25% expected by chance is statistically significant (z = 1.86, p = 0.031, one tail), supporting, therefore, the ESP hypothesis. The distribution of ranks of the target stimulus, displayed in table 1, reveals an interesting increasing pattern that provides further support to the ESP hypothesis. Sixty-two per cent of the individuals ranked the correct target either first or second, while only 38% did as their last two choices. This difference is highly significant (z = 2.52, p < .006, one tail).

Table 1. Expected and observed frequencies for each target rank.

	Ranks				
	1st	2nd	3rd	4th	Total
Expected	22.5	22.5	22.5	22.5	90
Observed	30	26	20	13	90

Bearing in mind randomicity criticisms made to previous ESP studies (see e. g. Hyman, 1985), target selection bias was tested for equi-probability of target, set number, and doublets inter-trial independence. Operating at an α level of .01, the distribution of targets for the 90 sessions proved to be random for the target alternatives (that is A, B, C, D; χ^2 = 4.75, n. s.) and set number (1 to 10; χ^2 = 4.59, n. s.), showing that all stimuli in the pool were selected as target with

statistically equal frequency. The position of the target stimulus and decoys in the judging sequence was also random (χ^2 = 2.94, n. s.), discarding, like this, the possibility that participants could have chosen the right stimulus due to positions preferences. Inter-trial independence was also tested for the 16 combinations of doublets (i. e. AA, AB, AC, AD, BA, BB, etc.), as the target could appear in each 2 consecutive sessions, showing further accuracy of the target distribution (χ^2 = 9.9, n. s.).

Variable assessment

Only four of the predictors explored in this study showed a significant association with the participants' ESP scores at an α=.01. These were internal awareness, self-reported feeling vigorous prior to the session, concern about the external environment during the Ganzfeld stimulation, and feeling upset during the experiment. In addition, seven further variables showed p-values below .05 (see table 2). Despite α-levels were adjusted to .01, as explained in the introduction, and keeping in mind that the correlations between these latter variables and the participants' ESP scores could have appeared by chance, we believe that it is worth taking these variables into account in our discussion. Thus, emotions like anger and frustration, tense arousal, feeling upset during the session, feeling uncomfortable, being concerned about the external environment during the Ganzfeld stimulation, as well as traits empathy and psychoticism, appeared positively associated to the participants' ESP scores. On the other hand, variables internal awareness, vulnerability, feeling vigorous, and sender-receiver intimate bond showed a negative association.

Table 2: Correlation coefficients and p-values (<.05) of ESP predictors.

Variable	Correlation Coefficient	Significance Level
Anger and frustration during the Ganzfeld	.18	.04
Tense arousal during the Ganzfeld	.17	.049
Feeling upset during the session	.27	.007
Feeling uncomfortable during the Ganzfeld	.21	.02
Sender-receiver intimate bond	-.19	.032
Empathy	.21	.024
Psychoticism	.18	.042
Concern about the external environment during the Ganzfeld	.26	.006
Internal awareness	-.31	.002
Vulnerability	-.24	.01
Feeling vigorous	-.26	.006

The positive association found between the individual's ESP scores and emotions of *anger/frustration*, *tense arousal*, *feeling upset*, and dis*comfort* during the Ganzfeld contradict the hypothesised relationships though only *feeling upset* does to a .01 significance level. These feelings and emotions were expected to have a detrimental effect on this test through a variety of mechanisms. For example, the individual's self-reported *tense arousal* and *discomfort* during the Ganzfeld stimulation could reflect a degree of anxiety that, according to mental state optimisation models from the literature (e.g., Braud, 1977; Honorton, 1977, 1978), would interfere with the mental and physical relaxation theoretically required for this type of task. However, the positive effect found for these variables could be understood if a certain level of arousal was necessary for successful performance, as suggested in some previous studies (e.g., Braud, 1977). The sample of participants could have been under aroused during the sensory monotonisation, for example due to the comfort of the reclining chair used in the experiments. Thus, those individuals who experienced tension and discomfort might have kept a level of activation more suitable to perform the task.

Psychoticism, empathy, and concern about the external environment during the Ganzfeld, as hypothesised, showed positive

associations with the participants' scores. As explained in the introduction, the concept of empathy, as the understanding of others' emotions, feelings and points of view, suggests that an ESP mechanisms could be involved in this process. Our results support this hypothesis. The association between psychoticism and ESP could be attributed to the components of non-conformity, disregard of and rebelliousness towards convention and social rules in this construct. It would be possible that individuals high on this trait found it easier to undertake a task conventionally viewed as impracticable and, even more, used it as an opportunity to aggress social convention. The relationship between the individuals' ESP scores and their reported concern with the external environment during the sensory monotonisation could have resulted from a successful attempt by the participants to focus on the relevant source of information (i. e. the sender).

On the basis of previous findings *Neuroticism* was expected to appear negatively associated to the participants' scores (e. g. Braud, 1977; Palmer, 1978; Weiner and Zingrone, 1986). None of the six traits forming this factor in the Big-Five Factors Model of Personality (Costa and McCrae, 1992) correlated significantly at an α = .01 with our participants' ESP scores. Only the trait vulnerability appeared in the expected direction with a p-value below .05. The negative association observed between variables *internal awareness, feeling vigorous,* and *sender-receiver intimate bond* and the participants' performance contradict our expectations. *Internal awareness* was expected to contribute to the ESP process in tune with signal detection theories. It was reasoned that participants who were more aware of their own thoughts, feelings, etc. would also be more likely to detect any ESP signal. However, it could also be the case that these individuals are, in addition, more aware of cognitions and mental imagery irrelevant to the target of the Ganzfeld task. Thus, the signal to noise ratio could not increase but decrease, explaining, like this, these participants' equal or lower scores. Variable *feeling vigorous,* in principle, would have been interpreted as a subjective indicator of good state of health. This was expected to contribute to the good performance of the experimental

task. However, this measure could also have reflected a degree anticipatory anxiety prior to the session and/or a difficulty to relax in the laboratory. This could result in physiological noise (as in the NRM) and, therefore, interfere with the individual's task of detecting the ESP signal. The expectations for higher scores by participants who were more closely related (friends and relatives acting as sender and receiver in the same session) were also based on previous literature findings (e.g., Bierman *et al*., 1993; Broughton *et al*., 1989) and characteristics of spontaneous case reports. However, the observed results did not confirm these expectations.

Other variables, like *impulsivity, absorption, altered state of consciousness, expectancies of success, phase of menstrual cycle in females, need for cognitive closure or controllability of mental imagery,* showed small, non-significant correlation indices. However, as explained in the introduction, there has been little research on the role of these variables in the ESP process and our hypotheses were perhaps too speculative. It was more striking to observe, similarly, that other variables that have previously appeared associated to ESP scores in a considerable number of studies, like *extraversion, prior psi laboratory testing, practice of mental disciplines,perceptual defensiveness, feeling and perception* of the Myer-Briggs model or *paranormal belief* did not seem to have an effect in this particular study. In the case of *creativity*, the lack of consensus in the definition of this construct in mainstream psychology and the subsequent disagreement about a standard measure (Feldhunsen and Goh, 1995) makes it very difficult indeed to draw safe conclusions about its relation to ESP.

Logistic Regression Analysis

A stepwise forward logistic regression analysis was performed on the eleven predictors that showed correlation coefficients with *p*-values of .05 or less. After four steps, a four variable solution was reached. The form of the equation is shown below. However, given that the original number of variables had to be narrowed down to a smaller sample consisting of only those with *p*-values under .05 before entering

them in the equation, there is room for the possibility that these observed relationships could reflect sampling error.

$$\text{PROB (hit)} = \frac{1}{1 + e^{-11.5 - .54(IA) - 1.02(FV) + 1.73(CE) - .21(VU) + \varepsilon}}$$

Note: where ε is the error term.

According to this model, the probability for a right guess {PROB (hit)} in a given experiment is accounted by four variables. In the equation, CE means "concern about the external environment" (during the period of Ganzfeld stimulation) and contributed to the session outcome with a positive coefficient in the equation of 1.73. FV, IA, and VU mean "feeling vigorous" (prior to the testing), "internal awareness", and "vulnerability" (from the neuroticism sub-scale of the NEO-PI), contributing negatively to the session outcome with coefficients –1.02, -.54, and -.21 respectively. (ε is the error term).

The equation classified correctly 79.55% of the cases, predicting accurately 91.5% of the misses and 55.17% of the hits vs. a baseline of 66% of misses and 33% of hits. A test of the model against the constant-only model was statistically significant, χ^2 (6, 88) = 33.4 ($p < .0001$), showing that the set of predictors was reliable in predicting the outcome of the session. The model accounted for 30% of the variance as indicated by the Hosmer & Lemeshow's goodness of fit statistic, R_L^2 (.30), an analogue of R^2 in multiple regression. Table 3 shows log likelihoods, Wald's statistics, partial correlations, step statistics, and percent of cases correctly classified for each step and predictor.

Table 3: Log likelihoods, Wald's statistics, partial correlation coefficients, step statistics, and percent of cases correctly classified for each step and predictor.

Step	Variable Entered	-2LL	Wald's Statistic	*r*	Step Statistic	Percent Correct (Hits)	Percent Correct (Misses)	Percent Correct (Total)
0	Constant	11.55	9.8***			0%	100%	67%
1	Internal awareness	103.2	7.48**	-.22	8.35**	34.48%	86.44%	69.32%
2	External concern	93.8	8.53**	.24	9.37**	37.93%	86.44%	70.45%
3	Feeling vigorous	83.23	8.85**	-.24	10.6***	55.17%	84.75%	75%
4	Vulnerability	78.15	4.48*	-.14	5.08*	55.17%	91.53%	79.55%

*$p < .05$, **$p < .01$, ***$p = .001$

A χ^2 test run on the remaining variables shows that addition of new variables will not increase the predictive power of the model (χ^2 = 14.7, p = .19). The amount of unexplained information that remains in the model is indicated by a –2 log likelihood statistic (–2LL) of 78.15 (a perfect solution would be associated to a –2LL of 0). The contribution of each predictor can be estimated from the loss/gain in log likelihood since last step when the term is added/removed. The Wald's statistic, used to test the significance of the β coefficient for each predictor, reaches statistical significance for all variables included in the equation. The step value, that indicates the improvement on the predictive power of the equation since the last stage, is equal to –2LL at the current step minus –2LL at the previous step. This can be taken as an indicator of the contribution of the predictor entered in the step to the predictive power of the model. The correlation between the observed values and the ones predicted by the equation, another indicator of the accuracy of the model, was also high and significant (r_{xy}=.71, $p < .001$, one tail).

Despite the large body of research conducted in the area, inconclusive results leave yet uncertain what parameters sustain a phenomenon that has been observed spontaneously in previous

studies. After a previous series of experiments, and in line with Bem, Palmer, and Broughton (2001) we concluded that the lack of replication of positive results observed in the literature could be a consequence of the heterogeneity of the experimental environments and procedures used across the different studies and recommended, based on our own results, the use of the Ganzfeld technique in combination with multisensory targets in the exploration of further predictors of ESP. The present study was a replication of these experimental conditions on a larger sample (N=90) and an exploration of those individual differences and mental states that appeared most successfully related to ESP performance in our previous study as well as in the literature.

The participants' rate of success in this study (33%) was lower than in our previous one (42%). However, the fact that, as in the previous study, it was significantly higher than chance expectation (25%) suggests that Ganzfeld stimulation together with multisensory targets are promising conditions for the testing of ESP. A logistic regression equation was obtained to predict the probability of a right guess on the basis of two stable traits (internal awareness and vulnerability), a pre-session state variable (feeling vigorous), and an on-session mental state indicator (concern about the external environment). It was curious to observe, however, that none of the variables which appeared most recurrently in previous research (i.e., extraversion, creativity, defensiveness, paranormal belief, feeling-perceiving, etc.) contributed to the final equation. This equation was, however, compounded of four variables barely considered before in the prediction of ESP scores. Researchers are encouraged to explore 'new' variables in future studies, though much work might be needed before we are able to outline a robust set of variables that warrants successful prediction, and replication, of a positive experimental outcome.

Acknowledgements

We gratefully acknowledge the financial support of Fundación Séneca, Centro de Coordinación de la Investigación in Murcia (Spain).

References

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

Bem, D.; Palmer, J.; and Broughton, R. (2001). Updating the Ganzfeld database: A victim of its own success? *Journal of Parapsychology*, 65 (3), 207-218.

Bierman, D. J.; Bosga D.; Gerding, H.; and Wezelman, R. (1993). Anomalous information access in the Ganzfeld: Utrecht-novice series I and II. *Proceedings of the Parapsychological Association 36th Annual Convention*, Toronto, Canada, p. 192-203.

Braud, W. G. (1975). Conscious vs. unconscious clairvoyance in the context of an academic examination. *Journal of Parapsychology,* 39(4), 277-288.

Braud, W. G. (1977). Psi conducive conditions: Explorations and interpretations. In B. Shapin and L. Coly (Eds.), *Psi and States of Awareness*, p. 221-237. New York: Parapsychology Foundation.

Braud, W. G. and Braud, L. W. (1973). Preliminary explorations of psi conducive states: Progressive muscular relaxation. *Journal of the American Society for Psychical Research,* 67(1), 26-46.

Broughton, R. S. (1977). An exploratory experiment on psi-based subjects and experimenter effects. In J. D. Morris, W. G. Roll and R. L. Morris (Eds.). *Research in Parapsychology, 1976*, p.173-77. Metuchen, NY: Scarecrow Press.

Broughton, R. S., Kanthamani, H. and Khilji, A. (1989). Assessing the PRL success model on an independent Ganzfeld base. In. L. Henkle and J. Palmer(Eds.), *Research in Parapsychology, 1989*, p.32-35. Metuchen, NJ: Scarecrow Press.

Costa, P. T. and McCrae, R. R. (1992b). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory NEO-FFI): Professional manual (Rev. Ed.). Odessa, FL: Psychological Assessment Resources.

Dalton, K. (1997). Is there a formula to success in the Ganzfeld? Observations on predictors of psi-Ganzfeld performance. *European Journal of Parapsychology,* 1997, 13, 71-82.

Dollard, J., Doob, L.W., Mowre, O.H. & Sears, R.R. (1939). *Frustration and Aggression.* New Haven, CT: Harvard University Press.

Eysenck, H.J. (1967). *The biological basis of personality.* Springfield, IL: Thomas.

Eysenck, H. J. and Eysenck, S. B. G. (1991). *Manual of the Impulsivity, Ventureness, and Empathy Questionnaire*. Hodder and Stoughton Educational. Kent (UK).

Feldhunsen, J. F. and Goh, B. E. (1995) Assessing and Accessing Creativity: An Integrative Review of Theory, Research, and Development. *Creativity Research Journal*, 8 (3), 231-247.

Fenigstein, A.; Scheiner, M.; and Bush, A. (1975). Public and private self-conciousness. Assessment and theory. *Journal of Consulting and Clinical Psych*ology, 43(4), 522-27.

George, L. (1982). Enhancement of psi functioning through mental imagery training. *Journal of Parapsychology,* 46(2), 111-125

Gordon, (1949). An investigation into some factors that favour the formation of stereotypical images. *British Journal of Psychology,* 39, 156-167.

Hansen, G., Schlitz, M., and Tart, C.T. (1984). Summary of remote viewing experiments. In R. Targ, & K. Harary, The Mind Race. New York: Villard Books.

Haraldsson, E. (1985). Interrogative suggestibility and its relationship with personality, perceptual defensiveness and extraordinary beliefs. *Personality and Individual Differences,* 6(6), 765-767.

Haraldsson, E. and Houtkooper, J. M. (1991). Psychic experiences in the multinational human values study: Who reports them? *Journal of the American Society for Psychical Research,* 85(2), 145-165.

Haraldsson, E. and Houtkooper, J. M. (1995). Meta-analyses of 10 experiments on perceptual defensiveness and ESP: ESP scoring patterns and experimenter and decline effects. *Journal of Parapsychology,* 59(3), 251-271.

Herrmann, N. (1999). *The Herrmann Brain Dominance Instrument*. Herrman International, Lake Lure (US).

Honorton, C. (1977). Psi and internal attention states. In B. Wolman (Ed.) *Handbook of Parapsychology*, p. 435-472. Jefferson, North Carolina: McFarland and Company.

Honorton, C. (1978). Psi and internal attention states: information retrieval in the Ganzfeld. In B. Shapin and L. Coly (Eds.), *Psi and States of Awareness*, p. 79-100. New York: Parapsychology Foundation.

Honorton, C. and Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology,* 53(4), 281-308.

Irwin, H. J. (1994). Paranormal belief and proneness to dissociation. *Psychological Reports,* 75(3, Pt 1), 1344-1346.

Kragh, U. (1960). The Defence Mechanism Test: A new method for diagnosis and personnel selection. *Journal of Applied Psychology, 44,* 303-309.

Kragh, U. (1985). Defence Mechanism Test. *DMT Manual*. Stockholm: Persona.

Kanthamani, B. K. and Rao, K. R. (1972). Personality characteristics of ESP subjects: III. Extraversion and ESP. *Journal of Parapsychology,* 6(3), 198-212.

Kanthamani, B. K. and Rao, K. R. (1973). Personality characteristics of ESP subjects: V. Graphic expansiveness and ESP. *Journal of Parapsychology,* 37(2), 119-129.

Kersey, D. (1978). *Please Understand Me*. Prometheus Nemesis Book Company. Del Mar, CA (US)

Kripper, S., Ullman, M., & Honorton, C. (1971). A precognitive dream study with a single subject. *Journal of the American Society for Psychical Research*, 65, 192-203.

Kruglanski, A. W. (1990). Lay episthemic theory in social cognitive psychology. *Psychological Inquiry*, 1 (3), 181-197.

Lawrence, T. (1993). Gathering in the sheep and goats: A meta-analyses of forced choice sheep/goat ESP studies, 1947 - 1993. *Proceedings of the Parapsychological Association 36th Annual Convention,* Toronto, Canada, p. 75-86.

Matthews, G.; Jones, D.M.; and Chamberlain, A.G. (1990) Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, 81, 17-42.

Milton, J. (1998). A meta-analysis of waking state of consciousness, free response ESP studies. In N. L. Zingrone, & M. J. Schlitz (Eds.). *Research in Parapsychology, 1993,* p.31-34. Lanham, MD: Scarecrow Press.

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

Milton, J., & Wiseman, R. (2001). Does psi exist? Reply to Storm and Ertel (2001) *Psychological Bulletin,* 127, 434-438.

Milton, J. and Wiseman, R. (2002) A response to Storm and Ertel (2002) *Journal of Parapsychology*, 62, 183-185.

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

Morris, R. L. and Morrell, N. M. (1985). Free-response ESP training: A significant failure. In R. A. White and J. Solfvin (Eds.) *Research in Parapsychology, 1984,* 5—54. Metuchen, NJ, Scarecrow Press.

Musso, J. R. and Granero, M. (1982). Group GESP experiments tending to yield repeated positive results. In W. G. Roll, R. L. Morris, and R. A. White (Eds.) Research in Parapsychology, 1981, p.100-103. Metuchen, NY: Scarecrow Press.

Neuberg, S.; Judice, T.; and West, S. (1997). What the need for closure measures and what it does not: Toward differenciating among related epistemic motives. *Journal of Personality and Social Psychology*, 72(6), 1396-1412.

O'Keefe, D. J. (2003). Against family-wise alpha adjustment. *Human Communication Research*, 29, 431-447.

Palmer, J. (1978). The out-of-body experience: A psychological theory. *Parapsychology Review,* 9(5), 19-22.

Palmer, J.; Khamashta, K.; and Israelson, K. (1979). An ESP Ganzfeld experiment with transcendental meditators. *Journal of the American Society for Psychical Research,* 73(4), 333-348

Parker, A.; Grams, D.; and Pettersson, C. (1998). Further variables relating to psi in the Ganzfeld. *Journal of Parapsychology,* 62(4), 319-337.

Pekala, R. J.; Kumar, V. K.; and Marcano, G. (1995). Anomalous/paranormal experiences, hypnotic susceptibility, and dissociation. *Journal of the American Society for Psychical Research*, 89, 313 - 332.

Pérez-Navarro, J. M. (2003). *Developing a recipe for success in ESP experimental research.* Unpublished doctoral dissertation. University of Coventry, Coventry.

Pérez-Navarro, J. M., Lawrence, T., and Hume, I. (2009). Personality, Mental State and Procedure in the Experimental Replication of ESP: A Preliminary Study of New Variables. *Journal of the Society for Psychical Research*, 73, 17-33.

Riley (1988). Measurement of dissociation. *The journal of nervous and mental disease*, 176 (7), 449-450.

Sargent, C. L. (1981). Extraversion and performance in "extra-sensory perception" tasks. *Personality and Individual Differences,* 2(2), 137-143.

Sargent, C. L. (1982). An unusually powerful response-bias effect. *Parapsychology Review,* 13(3), 8-10

Schmeidler, G. R. (1950). Some relations between Picture-Frustration ratings and ESP scores. *Journal of Personality,* 18, 331-343.

Schmeidler, G. R. (1954). Picture-Frustration Ratings And Esp Scores For Subjects Who Showed Moderate Annoyance At The Esp Task. *Journal of Parapsychology,* 18, 137-152.

Schmeidler, G. R. (1982). A possible commonality among gifted psychics. *Journal of the American Society for Psychical Research*, 76, 53-58.

Schmitt, M. and Stanford, R. (1978). Free-response ESP during Ganzfeld stimulation. The possible influence of menstrual cycle phase. *Journal of the American Society for Psychical Research*, , 72, 177-82.

Stanford, R. G. (1979).The influence of auditory Ganzfeld characteristics upon free-response ESP performance. *Journal of the American Society for Psychical Research,* 73(3), 253-272.

Stanford, R. G. and Angelini, R. F. (1984). Effects of noise and the trait of absorption on Ganzfeld ESP performance. *Journal of Parapsychology,* 48(2), 85-99.

Stanford, R. G. and Mayer, B. (1974). Relaxation as a psi-conducive state: A replication and exploration of parameters. *Journal of the American Society for Psychical Research,* 68(2), 182-191.

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(2), 107-114.

Tart, C. (2000). Fear of psychic phenomena. In Leskowitz (ED). *Transpersonal hypnosis: Gateway to body, mind, and spirit*. Boca Raton, FL. US. CRC Press.

Tellegen, A. and Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology,* 83(3), 268-277.

Thalbourne, M. A. and 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(832), 172-186.

Wallach, M. A. and Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. Holt, Rinehart and Winston, NY (USA).

Watt, C.A. (1992). Defensiveness and Psi: Problems and Prospects. In L.A. Henkel & G.R. Schmeidler (Eds) *Research in Parapsychology, 1990.* Metuchen, NJ: Scarecrow Press.

Watt, C. & Ravenscroft, J. (1999). Defensiveness, neuroticism and extrasensory perception. *Journal of the American Society for Psychical Research*, 93, 341-354.

Weiner, D. H. (1982). The medium of the message: Information-coding style in psi processes. *Parapsychology Review*, 13(5), 9-11.

Weiner, D. H. and Zingrone, N. L. (1986). The checker effect revisited. *Journal of Parapsychology*, 50(2), 85-121.

Wilhelm, K. and Parker, G. (1988). The development of a measure of intimate bonds. *Psychological Medicine*, 18(1), 225-234.

European Journal of Parapsychology
Volume 24.1, pages 93-110

ISSN: 0168-7263

Student Brief: Examining the Case for Dream Precognition

Eric Robinson
School of Psychology,
University of Birmingham, U.K.

Abstract

This paper reports a covert experimental investigation of dream precognition. 100 participants attended two laboratory sessions under the guise of a dreaming and personality study. In the first session, a questionnaire measuring variables previously suggested to predict experimental psi performance; dream recall (Sherwood & Roe, 2003) and Novelty Seeking (Bem, 2008) was completed. Participants were also instructed to keep a dream diary the night of session one. In session two (the following day) using their dream diary, participants were presented with two texts of scene descriptions and asked to rate which possessed the strongest similarities to the content of their dreams. The two texts actually described two video clips, one of which was then randomly selected and viewed by the participant (the 'target' video clip). If dreaming could be retro-causally influenced by future experiences, it was hypothesised participants would show a preference towards selecting the text describing the target video. Overall, no significant effect was observed. However, a significant relationship between rated degree of similarity (of dreaming and video description) and performance was observed and dream recall was also found to significantly predict performance. Novelty Seeking did not predict performance. Methodological issues of this approach are also discussed.

Correspondence details: Eric Robinson, School of Psychology, University of Birmingham, United Kingdom. Email: exr888@bham.ac.uk.

Introduction

In Rhine's early collections of reported ESP experience 75% of precognition case reports were recorded as having been experienced in the state of dreaming (Rhine, 1958). Subsequent research has found similar figures, confirming that the vast majority of reported precognitive phenomena appear to be experienced whilst dreaming (Steinkamp, 2000; Drewes, 2002). Previous research experimentally investigating the experience of psi during sleep has produced significant findings. The Maimonides dream ESP studies of the 1960's and 1970' produced strong evidence for the existence of psi phenomena (Sherwood & Roe, 2003). From the 450 trials, a hit rate of 63% was exhibited (based upon blind judge's data), whereby we would expect 50% by chance (Sherwood & Roe, 2003). Although only three of the Maimonides studies investigated dream precognition (Krippner, Ullman & Honorton, 1971; Krippner, Ullman & Honorton, 1972; 'Pilot Sessions', as cited in Sherwood & Roe, 2003), greater than chance performance and impressive effect sizes were observed. Thus, these early studies appeared to provide support to a psi explanation for the numerous reported cases of proposed dream precognition.

Since the Maimonides programme experimental studies of dream telepathy, clairvoyance and precognition have produced mixed results. Of these later studies only three have explicitly investigated dream precognition (Sherwood *et al*., 2002; Sargent & Harvey, 1982; McLaren & Sargent, 1982). The post Maimonides dream precognition studies have been the least successful in providing evidence in support of psi, all producing performance at chance expectation. However, there may be a number of valid explanations for the unsuccessful post Maimonides dream ESP studies, including methodological issues (see Sherwood & Roe, 2003) and the use of very small sample sizes (Sherwood *et al*., 2002 for example).

Additionally, understanding of dream precognition is extremely limited. This may be explained through (to current knowledge) there being only 6 published empirical studies explicitly investigating dream precognition. This area has been largely neglected. The majority of

precognition or retro-causality research has ignored the psi conductive state of dreaming, testing participants whilst awake (Radin, 2004; Steinkamp, Milton & Morris 1998; Bem, 2003). Although this line of enquiry has produced some promising findings, survey studies consistently suggest that the majority of precognition cases occur whilst dreaming (Drewes, 2002). Thus, the present study aims to investigate precognition 'in its natural settings' and awaken interest in the area that has been overlooked of late.

A recognised weakness associated with previous experimental psi research is the tendency for researchers to ask participants to consciously 'be psychic' and to judge the identity of targets in the laboratory (Bem & Honorton, 1994). As psi has been suggested to operate below levels of conscious awareness (Stanford, 1990), simply demanding participants to consciously use such abilities would appear to be self defeating. Therefore, disguising the aims of ESP studies and using covert measures of psi may achieve more promising results (see Bem, 2003; Radin, 2004 for examples of this). Due to these reasons, through disguising our study aims to participants, the present study attempted to elicit psi unknowingly through a covert measure.

Participants were recruited under the pretence of a study investigating "dreaming and personality". During the first session participants completed a short questionnaire and instructed to keep a dream diary. The next day participants read two texts describing scenes, selected which possessed the strongest similarity to their dreams and then rated degree of similarity. Unbeknown to participants, the two texts describe two video clips stored on a laboratory computer. Using random number generating (RNG) computer software, one of the two video clips would then be randomly selected and viewed by the participant.

Thus, the study attempts to retro-actively influence participants dreaming by the viewing of the video clip. If events can be precognitively viewed during dreams, then it was hypothesised that participants would be more likely to match their dream content to the description of the video they later watched. Although the design lacks awakening of participants during REM sleep, which may improve

dream recall, this diary style methodology adopted has proven to be a promising paradigm (Sherwood & Roe, 2003). Additionally, session two taking place the next day was based upon previous findings that reported length of time between precognitive experience and event occurrence is commonly 1-2 days (Orme, 1974).

Although the successful Maimonides trials mainly used static images, there is a consensus that video clips may be more effective as ESP targets, as they are of much greater similarity to real life events (Bem & Honorton, 1994). Thus, two videos were chosen as the stimulus for the study. There inclusion was due to both possessing 'information rich content' (both videos possess a number of stimuli – see appendix B for descriptions). The choice to utilise 'information rich' videos was made due to the distinct possibility that dreaming may be subtly influenced precognitively (e.g., the brief appearance of a bridge in a dream that is featured in a target video), rather than an exact match of dream content and target video.

The present study also recruited participants with no selection criteria regarding dream recall. Sherwood & Roe (2003) argue that for home based dream ESP investigation "this is crucial for studies that do not employ deliberate awakening" (Sherwood & Roe, 2003, pp.105). Yet, such claims appear to have little empirical backing. Therefore, the also study aimed to examine whether self reported dream recall predicts home based dream ESP performance.

Finally, the study also aimed to investigate whether the previously suggested personality characteristic of 'Novelty Seeking' is predictive of precognition. Irwin & Watt (2007) suggest that novel sensation seeking is a predictor of spontaneous ESP experience and experimental ESP performance. Recently, experimental precognition research has added support to this argument. Bem (2008) reports participants defined as high in Novelty Seeking having outperformed participants not meeting this criterion in a number of recent experiments (Bem, 2008). Thus, Novelty Seeking tendencies will also be measured in the present study, to analyse whether in line with previous research participants high in this trait exhibit psi performance.

Method

Design

The study was a mixed groups design. The dependent variable for the study was the ratio of participants that selected the text describing the randomly selected target video ahead of non target video text description (otherwise known as 'hit rate'). There were three independent variables in the study; participant dream recall score, participant Novelty Seeking scores and rated degree of similarity between dream content and text ('similarity').

Participants

100 participants (35 males, 65 females) were recruited from the University of Liverpool. Research assistants recruited participants from undergraduate psychology classes. On recruitment, participants were informed the study investigated 'dreaming and personality', consisted of two sessions on consecutive days and further information would be provided during the first session. No reward was provided for participation.

Materials

Video clips were randomly selected through the use of pseudo random number generating (RNG) software run on a Windows 98 PC. The software uses the "Math.random" method within the JavaScript programming language to make random selections.

Video clips: Each video clip was silent and lasted two minutes, 30 seconds. Video one was of a deserted, military manned town and video two; a coastal beach front.

Session one questionnaire: The questionnaire used in session 1 consisted of ten questions. Of the 10, only 4 questions were analysed for the purposes of the present study (2 questions for dream recall and

2 questions for Novelty Seeking). The remaining 6 consisted of 4 filler questions to disguise the aims of the study and two questions regarding belief in the paranormal (which were included for a separate undergraduate research project).

To reduce participation time demands, a short self devised measure was used to determine participant tendency to recall dreams. The measure consisted of two questions measured on a 5 point Likert Scale (5 = Strongly Agree > Strongly Disagree = 1): "I remember most of my dreams", and "I hardly ever dream" (reverse scoring).

This resulted in a score of 2 – 10 for dream recall. Participants scoring 8 or above were defined as having 'good' dream recall and participants scoring 9 or above were defined as having 'exceptional' dream recall.

A short two question measure using Likert Scale questions (5 = Strongly Agree > Strongly Disagree = 1) was included to measure Novelty Seeking tendencies. The scale is based upon Bem's (2008) reported measure. The two questions were: "In general, I am easily bored", and "Sometimes it's best to stick to what you know" (reverse scoring).

Session two questionnaire: This consisted of the two descriptive texts of the videos (see appendix B). The decision to use descriptions of the two videos, rather than to show both of the videos and then show the target video again was due to the consideration of interference. It was assumed that if dream precognition can take place, watching the decoy video may also influence dream content. Due to the visual nature of dreaming it was hoped that a written description may reduce such an effect. Although both passages were systematically written by the author and checked for accuracy several times, no formal independent ratings of similarity between video and text description were made prior to the study.

The session two questionnaire comprised of two questions. The first asked participants 'to select which of the two extracts was most similar to their dreaming content'. The second asked participants to

rate the degree of similarity between chosen extract and dream content using a VAS (anchors; 'not at all similar' – 'very similar').

Procedure

Participants were informed that they would be taking part in a study that would involve two consecutive day sessions, "investigating dreaming and personality". On recruitment, participants were led to a computer laboratory at the University of Liverpool. In the first session, participants were asked to fill out the session one questionnaire. Once completed, the experimenter informed the participant that they would be required to complete a dream diary that night. Participants were given a dream diary and were instructed, on awaking in the morning, to make a detailed record of all dreams they could remember. Participants were also asked to read the instructions included with the dream diary (see appendix A) before going to bed that night, to ensure the task was fully understood. A convenient time was then arranged to meet the next day at the laboratory for session two. Before leaving, the experimenter informed the participant the second session would involve another short questionnaire and watching a video clip.

On arrival for session two participants were asked to produce the completed dream diary. Participants were then instructed to describe their dreams (with the dream diary as a prompt) to the experimenter (experimenters were instructed to prompt with questions about the dreams in order to refresh memory). Participants were then asked to complete the session two questionnaire. When completing the questionnaire participants were unaware that the two texts were actually descriptions of the two videos stored on the computer lab.

The experimenter then used the RNG software to randomly select one of the two videos. The participant then watched the video (the 'target' video clip for the trial), before being debriefed and thanked for their time. Over the two sessions, the study lasted approximately 30 minutes.[1]

[1] It was planned that if a participant was unable to recall any dream mentation they would be instructed to select a text randomly and assign a similarity rating of 0. However, all participants were able to recall some details of their previous nights dreaming.

Covert nature of the study: five experimenters collected the data over a period of two to three months at the University of Liverpool, an institute that had no parapsychological research group or teaching interest prior to the present study. The data was collected by research assistants with no previous research reputation in the subject matter. The possibility of the covert nature of the task being compromised was therefore limited.

Results

A trial was classed as a hit if the participant selected the text describing the target video as being the more accurate match of their previous nights dream content. If participants selected the text describing the non target video, the trial was classed as a miss. By chance expectation we would expect an equal numbers of trials registering hits and misses.

Overall Sample

In the overall sample (n = 100) 52% of participants exhibited a hit and 48% a miss. A chi-squared analysis was run to determine if this difference was significant. The results of the analysis revealed the difference to be non significant ($x^2_{(1,\ 100)} = 0.16$, $p = .69$). Results indicating that a psi effect was not observed in the overall sample.

Similarity

To examine if rated similarity predicted trial success, logistic regression analysis was undertaken between similarity rating and participant performance. Analysis revealed a significant positive relationship between similarity score and participant hit rate ($\beta = 0.174$, $Exp(\beta) = 1.190$, $p = .04$). This result suggests that an increase of 1 unit (scale 0 – 10) on the similarity scale resulted in a 19% higher chance of participants registering a hit.

Dream Recall

A logistic regression analysis was also undertaken between dream recall scores and participant performance. As hypothesised, analysis revealed a positive relationship between dream recall scores and participant hit rate (β = 0.198, Exp(β) = 1.219, p = .07). This result suggests that an increase of 1 unit (scale 1 – 5) on the dream recall scale resulted in a 22% higher chance of participants registering a hit. However, this relationship only approached significance at the 5% level.

To further examine the relationship between dream recall and psi performance, performance of high scorers on the dream recall measure was analysed. Participants (n = 54) with good dream recall (≥ 8 on the scale) registered a hit rate of 59%. A chi-square analysis was run to determine if the hit rate was significantly above chance expectation. The results of the analysis, although approaching significance, were found to be non significant ($x^2_{(1, 54)}$ = 1.85, p = .17). Participants (n = 26) with exceptional dream recall (≥ 9) recorded a hit rate of 69%. A chi-square analysis was run to determine if this hit rate was significantly greater than chance expectation. The results of the analysis were statistically significant ($x^2_{(1, 26)}$ = 3.86, p = .05), with participants performing significantly above chance expectation.

Novelty Seeking

A logistic regression analysis was also undertaken between Novelty Seeking scores and participant performance. Analysis revealed a non significant relationship between Novelty Seeking and performance (β = 0.01, Exp(β) = 1.009, p = .94). Results suggesting that participants self reported Novelty Seeking tendencies did not predict psi performance.

Randomness (Post-Hoc)

As a pseudo-RNG software was used, possible selection biases were examined. No evidence was found to suggest any bias of video selection by RNG or participant preference bias in text selection. Furthermore, we found no evidence that hit rates differed dependent on video selection (i.e., one video did not produce a greater hit rate than the other).

Discussion

The primary outcome measure of 'total sample hit rate' did not produce a statistically significant effect, as a hit rate of 52% (chance expectation = 50%) was observed. This finding suggests that as an overall sample, participants were no more likely to select the target video ahead of the decoy. Examining why the primary measure in this study failed to produce evidence in support of the psi hypothesis is of importance, if future research is to promote a better understanding of dream ESP. A number of suggestions concerning the methodology used may provide some explanation.

If we consider that 46% of participants reported themselves as possessing poor dream recall, this is likely to have made the task difficult for a significant proportion of participants. As even by awakening, a proportion of participants are likely to have forgotten many details of their dreams. Additionally, unlike the REM awakening method which was used in the highly successful Maimonides studies (Sherwood & Roe, 2003), participants in the present study only recorded dreaming on natural awaking. Therefore, although dreams were recorded by all participants, the nature of home based psi research results in participants being more likely to recall dreams immediately prior to awaking. Due to these considerations it is likely that only a small proportion of participant dreams were recorded. If one is to accept the existence of precognition, it is plausible the phenomena may well have taken place, but both the limitations of

human memory and methodology used in this study may have resulted in it being less detectable.

Similarity

As hypothesised, a significant positive relationship between rated similarity (of dream content and video description) and performance was observed. This finding suggests that participants rated degree of similarity between dreaming content and video predicted performance. If dream precognition was taking place, we would expect participants achieving hits to have the highest similarity scores, as this suggests that the actual nature of the future stimuli (the video) was influencing dream content. However, simply concluding 'the trials that produced hits possessed the strongest similarities between dreaming content and video description' would be misleading. Similarity ratings for trials were made individually by participants and it is highly likely these ratings differ between participants. Therefore, for the above conclusion to be drawn, the entire data set would have had to be independently judged by the same person. Nevertheless, the hypothesised significant finding does suggest that the degree to which participants thought their dreaming and video description possessed similarity predicted precognitive performance.

Dream Recall

Participants classified as possessing good dream recall achieved a hit rate of 59% and although this was in the hypothesised direction and approached significance, this difference was not statistically significant. Additionally, participants classed as having exceptional recall achieved an even higher hit rate of 69% and this finding did prove to be significantly significant. To assess whether dream recall and performance were positively related a further regression analysis was also undertaken. Analysis confirmed a positive relationship approaching significance. Overall these findings appear to support Sherwood & Roe's (2003) assumption that screening participants for

dream recall prior to dream ESP experiments is of importance. The combined results of these analyses suggest that future home based ESP research should only recruit participants confident in their ability to recall dreaming content. Participants scoring low or mid range on this scale did not perform above chance expectation, whereas it was only higher scorers that tended to exhibit any significant performance. Thus, in the future it may not simply be a case of screening for poor dream recall, but only recruiting participants with self reported high dream recall.

A note of caution should be made when drawing conclusions from these findings. The measure was only a self report of behaviour. It does not tell us whether good recall of dreams on the night of the study predicted performance. Thus, purely concluding that high recallers performance was due to their ability to extensively recall dreaming occurring on the night of the study is not strictly accurate. Although the dream recall measure is likely to be strongly correlated to this, it is not a direct measure. Therefore, future research explicitly investigating whether it is the act of dream recall of the study night that explains greater psi performance is advised.

Novelty Seeking

The final hypothesis was that participants defined as high on the personality characteristic Novelty Seeking would exhibit evidence of a psi effect. The hypothesis was based upon earlier findings from both spontaneous case reports and the laboratory (Irwin & Watt, 2007; Bem, 2008). Yet, our analysis revealed no such effect. As the present study attempted to replicate earlier findings suggesting Novelty Seeking tendencies, the scale used was a similar version to that of Bem (2008). However, it is important to note that this is only a short two item version, more extensive and validated measures exist (see Goldberg, 1999), and recent forced choice psi experiments have found these predict precognitive performance (Luke *et al*., 2008). Thus, due to using a short scale, it is possible our ability to differentiate between Novelty Seeking tendencies in our sample may have influenced the results.

Considerations

As with all new approaches the study possessed some methodological limitations. Independent ratings of similarity between text descriptions and videos were not undertaken. Ideally a pilot sample to assess similarity prior to the study should have been used, to ensure both passages were similarly accurate descriptions (to there respective videos) and dissimilar to each other. In theory, one text description may have been marginally more similar to its video than the other. However, although the content of the two videos are clearly different and both passages were constructed with great care, empirical data supporting this is lacking and this flaw would need to be addressed in future work.

Only two videos were used as stimuli throughout the study and this raises some concern. Thorough analysis was unable to find evidence for RNG bias, participant selection bias or an interaction between the two (i.e., participants achieving a significant hit rate when one video was the target). Yet, using a larger pool of possible targets would be strongly advised in future work. As theoretically, a salient story in the news relating to one of the two videos may well influence participant selection. In addition, potential effect could become even stronger if data were to be collected over a short period for example.

A related weakness of our small target pool is the possibility of participants precognitively viewing the decoy stimuli, as although they would not visually see it, they did later read a detailed description. Interference of decoy stimuli is somewhat of an unavoidable problem if a researcher wants to directly use participant opinion to decide if a trial is to be classed as a hit or miss. In theory participants may exhibit psi (through viewing the 'wrong' future stimuli whilst dreaming), but subsequently register a miss. Anecdotally it is common for researchers to comment on how strong similarities between mentation and decoy stimuli in ESP research often take place. Typically in Ganzfeld studies this weakness is overcome (at least to some extent) by using a large pool of potential target stimuli. As only one decoy stimulus was viewed in this study, this potentially produces a greater likelihood of

the decoy stimuli interfering and causing a 'psi miss'. Although, the observed significant relationship indicating a higher similarity between dream mentation and chosen text description resulted in a greater likelihood of achieving a hit suggests this is unlikely to have occurred in any great frequency.

Consideration of how target and decoy stimuli are presented in similar research may also be relevant. The commonly used protocol in ganzfeld and dream ESP research is to show participants an exact copy of the target stimuli, plus decoys. This study presented participants only with descriptions of the decoy and target video, in an effort to reduce such an effect. The rationale behind this was an assumption that the usually highly visual nature of dreaming may less influenced by text in comparison to the visual nature of videos. Testing the validity of such an assumption may be of importance.

Future research examining the importance of the covert nature of the study may also be of importance. Although there is some evidence that 'asking participants to be psychic' may increase anxiety or arousal, and subsequently inhibit performance (Schmeidler, 1988), whether or not using covert measures in the investigation of ESP is important is largely unknown. Our significant results suggest it may be an area worthy of further interest.

In conclusion, although the primary outcome measure did not produce evidence for psi, yet the significant findings of the present study promote the continuation of the home based dream ESP methodology. Hopefully further research can provide understanding of the effects observed in this study and other dream based ESP research. Finally, as this method is relatively inexpensive and easy to adopt, it is ideal for both larger research studies and student projects.

Acknowledgements

The research presented in this paper represents work conducted during the author's B.Sc in Psychology (although independent of the degree itself). The author would like to thank Jodie Johnstone, Nicola Power, Joanna Atkinson and Hannah Daykin from the University of

Liverpool for their assistance in data collection. Additional acknowledgement is also made to the anonymous reviewers of the first edition of this paper for their helpful comments.

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.*

Bem, D. J. (2008). Feeling the future III: Additional experimental evidence for apparent retroactive influences on cognition and affect. *Proceedings of Presented Papers: The Parapsychological Association 51st Annual Convention, 24-32.*

Bem, D. J. & Honorton, C. (1994). Does psi exist? *Psychological Bulletin, 115,* 4-18.

Delanoy, D.L. (1989). Characteristics of successful free-response targets: Experimental findings and observations. In L. A. Hnekel & R. E. Berger, *Research in Parapsychology 1988*. Meutchen, NJ: Scarecrow Press.

Drewes, A. A. (2002). Dr. Louisa Rhine's letters revisited: The children. *Journal of Parapsychology, 66,* 343-370.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe,* (Vol. 7, pp. 7-28). Tilburg, Netherlands: Tilburg University Press.

Irwin, H. J., & Watt, C. A. (2007). *An Introduction to Parapsychology.* North Carolina: McFarland & Company.

Krippner, S., Ullman, M., & Honorton, C. (1971). A precognitive dream study with a single subject. *Journal of the American Society for Psychical Research, 65,* 192-203.

Krippner, S., Ullman, M., & Honorton, C. (1972). A second precognitive dream study with Malcolm Bessent. *Journal of the American Society for Psychical Research, 66,* 269-279.

Luke, D. P., Roe, C. A., & Davison, J. (2008). Testing for forced choice precognition using a hidden task: two replications. *Proceedings of Presented Papers: The Parapsychological Association 51st Annual Convention, 111-122.*

McLaren, I., & Sargent, C. L. (1982). Awareness of success in free-response dream ESP testing with a single subject. In W. G. Roll, R. L. Morris, & R. A. White, *Research in Parapsychology 1981.* Metuchen, NJ: Scarecrow Press.

Orme, J. E. (1974). Precognition and time. *Journal of the Society for Psychical Research, 47,* 351-365.

Radin, D. (2004). Electrodermal presentiments of future emotions. *Journal of Scientific Exploration,* 18, 253-273.

Rhine, J. B. (1954). Frequency of types of experience in spontaneous precognition. *Journal of Parapsychology, 15,* 164-191.

Sargent, C. L., Harley, T. A. (1982). Precognition testing with free-response techniques in the ganzfeld and the dream state. *European Journal of Parapsychology, 4,* 243-56.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and mismatches.* Jefferson, NC: McFarland.

Sherwood, S. J., & Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP programme. *Journal of Consciousness Studies, 10,* 85-109.

Sherwood, S. J., Roe, C. A., Simmonds, C. A., & Biles, C. (2002). An exploratory investigation of dream precognition using consensus voting judging and static targets. *Journal of the Society for Psychical Research, 66,* 22-28.

Stanford, R. G. (1990). An experimentally testable model for spontaneous psi events: A review of related evidence and concepts from parapsychology and other sciences. In S. Krippner (Ed.). *Advances in Parapsychological Research Vol. 6.* NC: McFarland.

Steinkamp, F., Milton, J., & Morris, R. L. (1998). A meta-analysis of forced-choice experiments comparing clairvoyance and precognition. *Journal of Parapsychology, 62,* 239-246.

Steinkamp, F. (2000). Acting on the future: A survey of precognitive experiences. *Journal of the American Society for Psychical Research, 94,* 37-59.

Appendix A: Dream Diary Instructions

In this section, please describe all your dreams during the set night. Please also include dreams you may experience as you are drifting off to, and awaking from sleep. Descriptions should be as vivid as possible, if possible include setting, what was going on, who or what was there and running themes etc. Please include all information no matter how obscure or small in detail you feel it may be. This information will be kept confidential so do not worry about embarrassing or odd content, the majority of dreams are commonly of an unusual nature. Please use an additional sheet if you run out of the space required below.

Appendix B: Video Descriptions

Video one

This scene is of a seaside pier. The camera is focused on a sign declaring 'Danger of Falling debris Keep Out'. The camera then pulls out to reveal a large seaside pier. Wooden Stands support the structure, raising the pier several feet off the beach and sea floor. Several white and blue buildings are on the pier. Two turrets are noticeable at the end of the pier and two rounded building tops at the start of the pier. The camera then moves outwards to reveal a glistening sea, with the setting sunlight reflecting off the surface. The camera then moves to the right to reveal the beach front, waves gently crashing onto several long thin wooden and concrete structures on the beach, there to defend against sand loss. Slightly further on is the sea front, a pedestrian can be seen in the distance, walking along the road side. Several large white buildings can be seen in the distance. The camera then moves back to focus on the pier. Moving past the pier, the other side of the sea front can now be seen. Several large white buildings are again seen, as can large posts close to the beach. In the distance is a hillside, on the hill appears to be the remains of an old castle of some sort. The camera pulls out to reveal a pedestrian walking alongside the road close to the pier. Finally the camera focuses on the pier, revealing its entire length. This is a video of Hastings Pier, the castle remains are of Hastings Castle, all close to the beach and sea front.

Video two

This scene reveals two men, dressed in army style camouflaged combat attire, brushing a concrete forecourt outside. Close to them are several military looking vehicles also camouflaged, two traffic cones and a tree in the background. The camera then rotates to reveal another man brushing the forecourt; close to him are

several large industrial like containers. In the background is a tree and what appears to be a telephone poll leaning slightly to its side. The scene then changes to another military personnel, holding a gun and wearing a facemask. In the background appears to be a rail track and several other people slightly further on. The scene again changes, this time several troops are viewed close to a military truck on a road close to a bridge. Street lamps are at the side of the road. The camera then reveals a baron, desert like road side with some plantation, a telephone mast in the background and a bridge further on. A military vehicle drives past the camera. The scene then cuts to the outside entrance of a military hospital entrance, the outside of building is yellow brick, tangled wires are visible above the entrance. Two soldiers play fight. An African American soldier with a goatee beard and full battle attire then walks closely in front of the camera view. Another soldier is seen struggling to put on a glove. Four soldiers are then seen talking at the entrance, the sign on the entrance is 'C.C.P 4'. This is a video of soldiers in a deserted town during the Iraq war.

European Journal of Parapsychology
Volume 24.1, pages 111-114

ISSN: 0168-7263

Book Review

Etzel Cardeña
Center for Research on Consciousness and Anomalous Psychology (CERCAP),
Department of Psychology, Lund University, Sweden

A review of "Voyage to the Rainbow: Reminiscences of a Parapsychologist" by Milan Ryzl (2007)

Milan Ryzl is a Czech biochemist once associated with Rhine's Institute of Parapsychology and best known in our field for his work on using hypnosis to stimulate psi performance (e.g., Ryzl, 1966), his account of psi research in communist countries (e.g., Ryzl, 1968), and his description of a small effect, stable psi performance by Pavel Stepanek (e.g., Ryzl & Pratt, 1963), who seemed to lose his ability later on (Ryzl & Beloff, 1965). Because hypnosis has been linked to psi throughout history (e.g., Dingwall, 1967-68) and there continues to be some, although not unambiguous, theoretical and empirical support for this link (e.g., Cardeña, in press; Tressoldi & Del Petre, 2007), I was especially looking forward to reading this book.

Voyage to the rainbow, a self-published book, is the rambling auto-biographical account of Ryzl various life events and, even more so, his opinions about most everything under the Sun, including "human nature," religion, G. W. Bush, and individuals and traffic police from different nations. When Ryzl writes about parapsychological issues, he is usually more chatty than thorough, even in the second half of the book, which is devoted to his research and thoughts on psi. It is telling that the book lacks references other than those in a few footnotes and that he makes a number of elementary, factual mistakes, such as misunderstanding the meaning of depersonalization and confusing the contemplative despondency known as the "dark night of the soul" with a state of consciousness without an object.

Correspondence details: Etzel Cardeña, Thorsen Professor of Psychology, Department of Psychology, Lund, University, P. O. Box 213 SE-221 00 Lund, Sweden. Email: Etzel.Cardena@psychology.lu.se.

Voyage to the rainbow mostly covers three parapsychological areas, albeit superficially in my view. The first is his experience with fraudulent and potentially real psychics behind the Iron Curtain, including one who apparently provided him with important personal information. The second concerns what he considers one of his main discoveries, "the recognition of the relationship between ESP and religions (p. 167)." Using quotation from Christian, Hindu, Buddhist, and other traditions, he proposes that ESP refers to an immaterial aspect of creation, which he equates with God. Although I agree with some of the passages he quotes, they are not discussed in depth, as compared with a book like Huxley's *The Perennial Philosophy* (1990/1945). He bases his conclusions partly in research in which he asked talented participants to "use their ESP to obtain information about various topics of religious significance (p. 243). He mentions that participants asked these questions were studied independently from each other, but shows no awareness about the potential suggestive impact of his involvement in the hypnotic process.

The third parapsychology topic revolves around the association of psi and hypnosis. Ryzl's training on hypnosis was mostly or totally from buying popular books, as far as I could tell, and it shows. He states, for instance, that hypnosis is "a state similar to sleep (p. 187)", which is not the case, and he does not show any awareness that a hypnotic effect depends on hypnotizabilty, a trait-like ability to respond to hypnotic suggestions and have anomalous experiences, which manifests important individual differences (e.g., Cardeña et al., 2007). Ryzl's definition of the hypnotic "trance" at time sounds like strong absorption, at times like "consciousness without an object," but at no point did I find evidence that he had reviewed the literature on the area. Unfortunately he also seems to be unaware of much research in psi research other than the older literature and Stanford's PMIR theory (e.g., 1977), and he makes claims that must be qualified, for instance that psi-missing occurs when the participant is in an emotionally negative state (while there may be a relationship in this regard, it is far from robust, and may interact with other factors such as the sheep-goat effect, which he does not address). Also, I found his

statements about how hypnosis can provide the *via regia* to the development of psi abilities overblown, unless he has evidence (which was not evident in the book) that not only Stepanek, but that he himself and many others can achieve strong and demonstrable psi abilities just by using hypnosis. The cumulative literature on psi makes difficult to escape the conclusion that there are striking individual differences in people's abilities to evidence psi, at least in controlled experiments, although it is also clear that we understand very little about what those abilities are.

I did find, however, one chapter that I think is worth rescuing from the book, the appendix on "mental impregnation." In it, Ryzl proposes that psychical attention or intention can "impregnate" matter with information. For instance, he reviews his data on his participant Stepanek, who tended to give the same description as he had previously given (either green or white) to a card that was placed in different envelopes at different times. Unless there was some kind of physical leak, this effect indeed seems difficult to explain in ordinary terms and is consistent with Roll's (1982) theory on hauntings and similar phenomena.

As for the sections on Ryzl's views about the world, they often display various prejudices, some of which I found quite distasteful. For instance, his opinion that poverty in major cities can be explained by lazy country people coming to the city to beg instead of work, or his conclusion that a Brazilian street boy "is winning over the system (p. 137)" when he got some bread from Ryzl's table. He should live in the shoes (an obvious metaphor because probably the kid is barefoot) of that kid for 24 hours to see whether he is beating any system! I felt also that the section on Ryzl's judgment about and legal entanglements with various members of his family was unfortunate and completely gratuitous. Although I do not know him or his family, I suspect his claims that he has "*always* [my emphasis] tried responsibly to do good, and to make other people happy (p. 200)." None of us is that perfect, psi or no psi… In sum, I do not recommend embarking on Milan Ryzl's voyage…

Publication Details: *Milan Ryzl. (2007). Voyage to the Rainbow: Reminiscences of a Parapsychologist. Victoria, Canada: Trafford Publishing. 274 pp. ISBN 1-4251-1233-1. Publication price: US$22.95, EUR17.89, £11.86. (Paperback).*

References

Cardeña, E. (in press). Anomalous experiences during deep hypnosis. In M. Smith (Ed.) *Developing perspectives on anomalous experience*.

Cardeña, E., Lehmann, D., Jönsson, P, Terhune, D., & Faber, P. (2007). The neurophenomenology of hypnosis. Proceedings of the 50th Annual Convention of the Parapsychological Association, 17-30.

Dingwall, E. J. (Ed.) (1967-8). *Abnormal hypnotic phenomena: A survey of nineteenth-century cases,* 4 Vols. London: J, & A. Churchill.

Huxley, A. (1990). *The perennial philosophy.* New York: Harper Perennial. (Originally published 1945).

Roll, W. G. (1982). The changing perspective on life after death. In S. Krippner (Ed.), *Advances in parapsychological research 3* (pp. 147-291). New York and London: Plenum

Ryzl, M. (1966). A method of training in ESP. *International Journal of Parapsychology, 8,* 501-532.

Ryzl, M. (1968). Parapsychology in communist countries of Europe. *International Journal of Parapsychology, 10,* 263-276.

Ryzl, M., & Beloff, J. (1965). Loss of stability of ESP performance in a high-scoring subject. *Journal of Parapsychology*, *29*,1-11.

Ryzl, M., & Pratt, J. G. (1963). A further confirmation of stabilized ESP performance in a selected subject. *Journal of Parapsychology, 27*, 73-83.

Stanford, R. G. (1977). Conceptual frameworks of contemporary psi research. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 823-858). New York: Van Nostrand Reinhold.

Tressoldi, P. & Del Prete, G. (2007). ESP under hypnosis: The role of induction instructions and personality characteristics. *Journal of Parapsychology, 71,* 125-137.