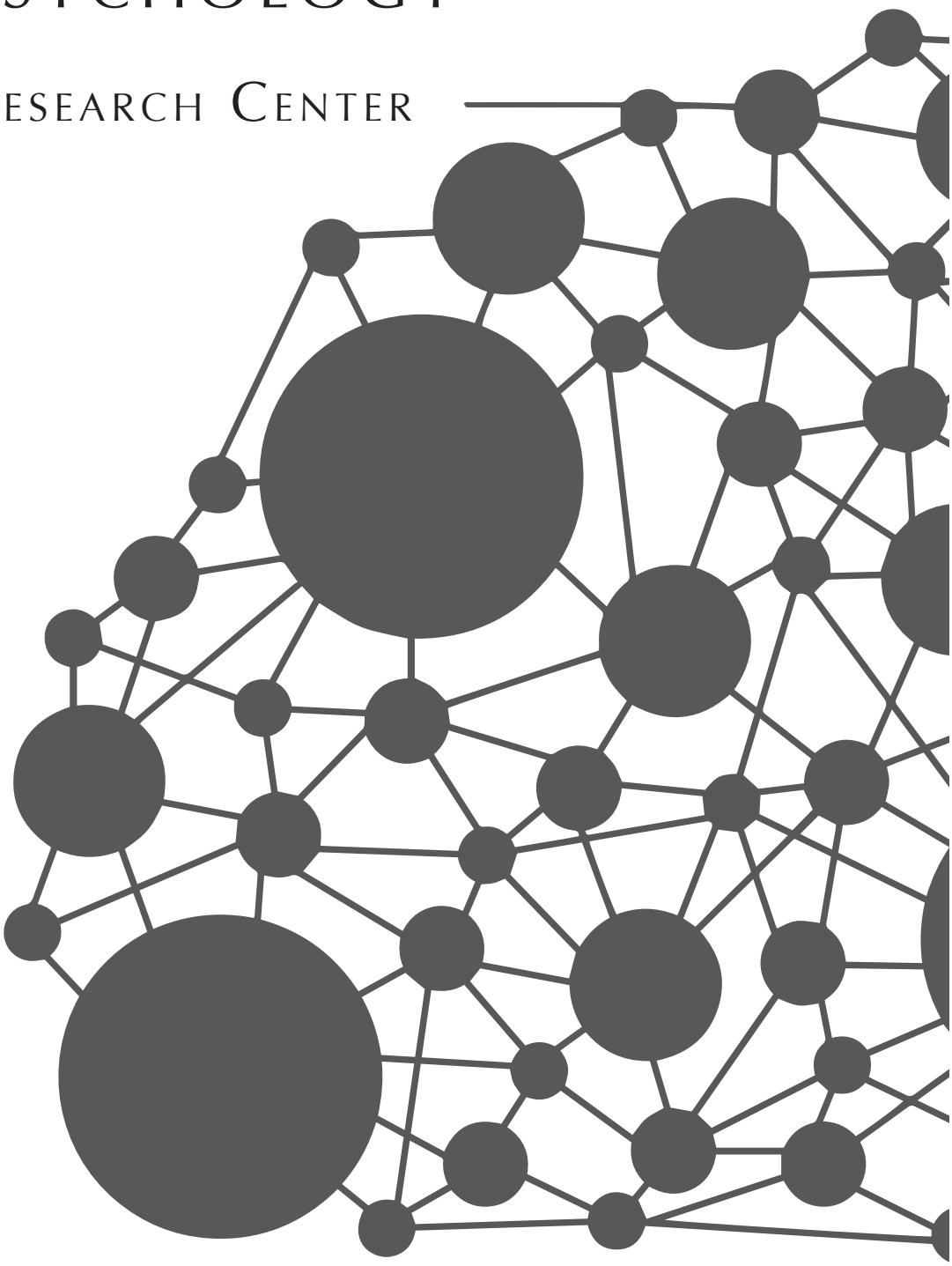
Journal of Parapsychology

— Rhine Research Center



Volume 81, Number 1 Spring 2017 ISSN 0022-3387

The *Journal of Parapsychology* is published twice a year, in Spring and Fall, by Parapsychology Press, a subsidiary of The Rhine Center, 2741 Campus Walk Ave., Building 500, Durham, NC 27705. The *Journal* is devoted mainly to original reports of experimental research in parapsychology. It also publishes research reviews, methodological, theoretical, and historical papers of relevance to psi research, abstracts, and selected invited addresses from Parapsychological Association conventions, book reviews, and letters.

An electronic version of the *Journal* is available to all subscribers on the Rhine Research Center's website (www.rhine.org.) The current subscription rates are: Individuals ($65.00), institutions ($77.00), with no other categories available. Members of the Rhine Research Center in the Scientific Supporter category receive the electronic journal free with their membership. The current subscription rates for paper copies of the *Journal* are: Individuals ($100.00), institutions ($118.00). Foreign subscribers must pay in U.S. dollars. Selected single issues (current or archival) are available at $35.00 each; go to www.rhine.org for more information. Orders for subscriptions or back issues, correspondence, and changes of address should be sent to: *Journal of Parapsychology*, 2741 Campus Walk Ave., Building 500, Durham, NC 27705. Subscriptions may also be ordered online at www.rhine.org.

Postmaster: Send address changes to the Journal of Parapsychology, 2741 Campus Walk Ave., Building 500, Durham, NC 27705. Subscribers: Send change of address notice 30 days prior to the actual change of address. The Journal will not replace undelivered copies resulting from address changes; copies will be forwarded only if subscribers notify their local post office in writing that they will guarantee second-class forwarding postage.

The *Journal* is an affiliated publication of the Parapsychological Association.

ISSN 0022-3387

Volume 81 / Number 1 / Spring 2017

EDITORIAL

ARTICLES

TECHNICAL NOTE

BOOK REVIEWS

We thank the following for their translations of the abstracts: *Eberhard Bauer* (German), *Etzel Cardeña* (Spanish), and *Renaud Evrard* (French)

Journal of Parapsychology, 81(1), 6-8

EDITORIAL

WHITHER PARAPSYCHOLOGY?

By Etzel Cardeña

In a supplement to this journal celebrating its 75th anniversary, various authors gave their sense of where they think the field will be in 25 years (Palmer, 2012). Most contributors were optimistic but also somewhat uncertain as to the exact shape of parapsychology in the future, and with good reason. The scientific study of parapsychological (psi) phenomena has never been an easy task, surviving, sometimes barely, a perfect storm of opposing forces: the seeming fitfulness of the phenomena, an organized and sometimes nasty movement to prohibit their mere study, and, on the other extreme, the uncritical acceptance of various evidence-free and self-contradictory beliefs (Cardeña, 2011, 2015b). During my lifetime I have witnessed cycles in which parapsychology seemed to be on the verge of becoming accepted by mainstream science to then fall and seem to be gasping for air.

I accepted the kind invitations of the previous *Journal of Parapsychology* Editor, John Palmer, and of John Kruth, Executive Director of the Rhine Research Center, to become the new editor of the *Journal,* not without some trepidation. I have always had the good fortune (others might describe it as a curse) to be interested in areas that are complex and fascinating, including the implications of hypnotic phenomena for human potentials, and pathological and non-pathological anomalous experiences and alterations of consciousness. Recognition of the intrinsic value of these areas to our understanding of the mind has been hard-fought and hard-won, but the case for psi phenomena remains in contention, at least as far as an open embrace by mainstream science. I wrote "open" advisedly because I, as well as other researchers in the field, have met many scientists who usually remain mute about the topic but then approach me and tell me in a hush-hush tone that they have personally experienced a psi event. So where does parapsychology stand currently?

There are various supportive indicators including:

a) publications in top-ranked journals of various disciplines (e.g., Bem, 2011. Radin, Michel, & Delorme, 2016; Storm, Tressoldi, & Di Risio, 2010),

b) various meta-analyses supporting the psi hypothesis (Cardeña, Palmer, & Marcusson-Clavertz, 2015),

c) the increasing academic recognition of parapsychology and related topics in the UK (Luke, 2012) and perhaps in other places,

d) the rediscovery of anomalous experiences, which have been associated with psi phenomena, as a legitimate area of scientific inquiry (Cardeña, Lynn, & Krippner, 2017),

e) the endorsement of continued psi research by mainstream past and current eminent scientists in various fields (Cardeña, 2014, 2015a),

f) a general interest in various aspects of psi represented by close to 1,500 publications in various disciplines, as evidenced in bibliographies prepared by Gerd Hövelmann for close to 10 years for the bulletin of the Parapsychological Association *Mindfield*.

But other indicators should temper any undue optimism. As compared with other topics, psi research suffers from:

a) few incoming or established researchers and centers of research throughout the world,

b) very little research funding and academic recognition,

c) a dispersal of the limited input of publications into new journals, some of them with questionable quality control,

d) an organized opposition to any support for psi, most obvious in wikipedia entries, and a more general uninformed bias against psi in other publications (Roe, 2015).

And then there is the mixed blessing, brought about to a large degree by the innovative work of Daryl Bem (2011), in which reasonable commentators remarked that his work was every bit as good as that accepted in other areas of psychology to then conclude that this *must* imply that there is something wrong in how research in other areas is conducted, evading the considerable support for psi in the research by Bem and many others.

As the incoming editor (John Palmer was the editor for all other items in this issue), I have a number of aims for my tenure:

1) "Mainstream" more the study of psi research by having a multidisciplinary editorial board holding different but informed views on the phenomena.

2) Expand the remit of the *Journal of Parapsychology* to include related areas such as anomalous experiences and states of consciousness in general, as the founders of the Society for Psychical Research had in mind for the then budding field.

3) Open the *Journal* even more to other research approaches than experimental, quantitative ones, since any method has particular strengths and weaknesses, which can be supplemented by other approaches.

4) Encourage contributions from various disciplines, including those not often represented in the *Journal* such as anthropology and the humanities.

5) Expand its international outreach by including reviews of books published in other languages than English.

6) Bring the journal up to date to the standards of other academic journals by, for instance, using digital object identifier system, or dois, and exploring the possibility of providing open access options to authors.

7) And, of course, maintain the level achieved by the *Journal* throughout its nearly 80 years of existence as the most important forum for psi research.

If the *Journal* can meet these goals, I hope that it will also gain a broader hearing by the scientific community at large, although I do not expect that anti-psi zealots of the past or the present will be persuaded. It has been very instructive for me to find out from journal reviewers and my own students that the psi community should reply when necessary but not focus on these critics because the silent majority does not hold their views. To accomplish all of these goals I will require the contribution of some of you who are reading these lines and will send your papers to the *Journal.* If the next editor can take for granted my list of wishes and then add some other ones, I will consider my (our) work done.

References

Bem, D. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100,* 407-425.

Cardeña, E. (2011). On wolverines and epistemological totalitarianism. (Guest editorial). *Journal of Scientific Exploration, 25,* 539-551.

Cardeña, E. (2014a). A call for an open, informed, study of all aspects of consciousness. *Frontiers in Human Neuroscience,* doi: 10.3389/fnhum.2014.00017.

Cardeña, E. (2015a). Eminent people interested in psi. *http://psi-encyclopedia.spr.ac.uk/articles/eminent-people-interested-ps.*

Cardeña, E. (2015b). The unbearable fear of psi: On scientific censorship in the 21st century. *Journal of Scientific Exploration*. 29, 601-620.

Cardeña, E., Lynn, S. J., & Krippner, S. (2017). The psychology of anomalous experience: A rediscovery. *Psychology of Consciousness: Theory, Research, and Practice, 4,* 4-22.

Cardeña, E., Palmer, J., & Marcusson-Clavertz, D. (2015). *Parapsychology: A handbook for the 21st century.* Jefferson, NC: McFarland.

Luke, D. (2012). Parapsychology & anomalistic psychology in the UK: A brief review. Mindfield, *4*(2), 59-63.

Palmer, J. (Ed.) (2012). Where will parapsychology be in the next 25 years? Predictions and prescriptions by 32 leading parapsychologists. *Journal of Parapsychology, 76* Supplement. Retrievable from http://www.rhine.org/images/jp/jpsupplement122012.pdf

Radin, D., Michel, L., & Delorme, A. (2016). Psychophysical modulation of fringe visibility a distant double-slit optical system. *Physics Essays, 29,* 14-22.

Roe, C. (2015). What are psychology students told about the current status of parapsychology? *Mindfield, 7,* 86-91.

Storm, L., Tressoldi, P. E., & Di Risio (2010). Meta-analysis of free-response studies, 1992-2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin, 136,* 471-485.

Journal of Parapsychology, 81(1), 9-32

INDIVIDUAL DIFFERENCE CORRELATES OF PSI PERFORMANCE IN FORCED-CHOICE PRECOGNITION EXPERIMENTS: A META-ANALYSIS (1945–2016)[1]

By Marco Zdrenka and Marc Stewart Wilson

ABSTRACT: Previous research in parapsychology has not been particularly persuasive, in large part due to a lack of replicability of significant findings. To address these concerns and better understand which factors may be associated with stronger and more consistent effect sizes, all forced-choice precognition experiments analysing individual differences (e.g., personality traits) were aggregated to determine which factors might reliably predict psi performance. Overall, 55 studies published between 1945 and 2016, including 35 individual difference measures, were subject to meta-analysis. Six individual difference measures, namely, luck belief (the belief that luck is primarily controllable), perceptual defensiveness, openness to experience, belief in psi, extraversion, and time belief as dynamic, were found to significantly correlate with psi performance. Given the particularly straightforward nature of forced-choice precognition experiments, a promising future avenue would be to explore these factors in confirmatory studies. It is hoped that researchers can model their future experiments off these findings in conjunction with preregistration techniques, to ultimately create a more systematic and robust database.

Keywords: meta-analysis, psi, precognition, personality, individual differences, forced-choice

Statistician Jessica Utts (1991) made the statement that a "promising direction for future process-oriented research [in parapsychology] is to examine the causes of individual differences in psychic functioning" (p. 377). It seems that such an approach is not only reasonable but also necessary, given that the evidence for psi (a general term used to describe anomalistic communication or interaction with the environment) is often inconsistent and elusive (Kennedy, 2001). If psi is to be taken seriously by the scientific community, its nature needs to be observable under prespecified conditions (Alcock, 2003; Hyman, 2010). Individual difference factors—such as specific personality traits (e.g., extraversion) or beliefs (e.g., belief in psi)—have been extensively analysed and thus represent a promising avenue in this regard. However, many researchers ignore individual difference factors (potentially missing important sources of between-individual variation in psi performance) or look at a multitude of varied factors that are difficult to sort through. An actual effect may also be masked if an individual difference factor is systematically related to psi performance. For example, participants who score high in a trait may overperform while participants who score low in that trait may underperform, effectively cancelling each other out. Therefore, this meta-analysis was intended to synthesise the relevant research to better understand the factors that may lead to a successful (or unsuccessful) and consistent demonstration of psi in the laboratory.

This meta-analysis focuses specifically on forced-choice experiments that have tested for precognition (i.e., the foreknowledge of an event without any known explanation). Forced-choice experiments give participants several options to choose from for their response, whereas free-choice experiments allow participants to make an unrestricted response. As free-choice experiments—such as the ganzfeld—have received a lot of attention in recent literature (see Bem & Honorton, 1994; Milton & Wiseman, 1999; Storm, 2006), this meta-analysis focuses exclusively on forced-choice experiments. It also focuses on precognition rather than telepathy (anomalous communication between people) or clairvoyance (perception without using normal sensory modalities), as precognition experiments are less susceptible to sensory leakage (Steinkamp, 2005). For example, in some telepathy experiments, participants may potentially make decisions based on the sender's or experimenter's facial cues, but this is not possible in precognition experiments in which the

[1] An earlier version of this paper was presented at the 10th Biennial European Conference of the Society for Scientific Exploration in Sigtuna, Sweden, 2016.

target cannot be known (even by the experimenter) until after the participant has already made their choice.

Previously, there have been two meta-analyses conducted on individual differences in psi laboratory research: one looking at extraversion (Honorton, Ferrari, & Bem, 1998) and the other looking at belief in psi, or what is known as the sheep-goat effect (Lawrence, 1993). Both meta-analyses found a relationship with psi performance (r = .09 and r = .03, respectively), indicating a small yet robustly significant overall effect size. However, many other individual differences have been examined in individual studies and it would be useful to summarise those studies here, and to compare them all with one another. Furthermore, many meta-analyses combine studies from multiple domains, making it difficult to unpack exactly what factors constitute a replicable psi experiment. Thus, the purpose of this meta-analysis is to (a) provide a comprehensive and updated review of all forced-choice precognition experiments that have included individual difference measures, and (b) estimate the overall magnitude of the relationship between each individual difference measure and psi performance, with the overall goal to provide researchers with the necessary information needed to design confirmatory studies.

Meta-Analysis

Although some meta-analyses have focused on only one research paradigm, such as the ganzfeld (Hyman, 1985) or biological systems (Schmidt, Schneider, Utts, & Walach, 2004), this is the first meta-analysis to combine both an experimental paradigm and individual differences. Forced-choice precognition is the chosen paradigm, as it is the most efficient method available for replication; the experiments are often automated (less potential for interference from both participants and experimenters); and exact probabilities of hits/misses can be objectively calculated. Although free-choice experiments can also be quantified, this requires an additional step, as participant responses need to be converted to target responses. This is avoided in forced-choice experiments altogether.

Although forced-choice precognition experiments might seem too narrow a subset to analyse, Steinkamp (2005) reviewed all forced-choice extrasensory perception (ESP) experiments—including telepathy, clairvoyance, and precognition—and found it difficult to come to any conclusions due to conflicting outcomes and wide variations in study designs. Furthermore, whereas some studies do not show any differences in effect sizes between precognition and other domains (see Steinkamp, Milton, & Morris, 1998), other studies have found a difference between clairvoyance, precognition, and telepathy effect sizes (Storm, Tressoldi, & Di Risio, 2012; Tart, 1983). Therefore, in defining the inclusion criteria narrowly, we sought to overcome the heterogeneity of studies in Steinkamp's (2005) review.

For the purposes of this meta-analysis, the effect size of interest is the correlation coefficient between the individual difference measure and psi performance—not psi performance specifically—with the participant as the unit of analysis.

Method

Retrieval of Studies

Only studies in the published literature are included in the meta-analysis, since parapsychology is a relatively small field and it is unlikely that there are many unpublished dissertations or theses (Honorton & Ferrari, 1989). Sourcing of relevant studies included the bibliography of two meta-analyses (Honorton & Ferrari, 1989; Storm et al., 2012), a database search (described below), along with an inspection of all English-language parapsychological journals, namely, the *Journal of Parapsychology*, *Journal of the American Society for Psychical Research*, *Journal of the Society for Psychical Research*, *Research in Parapsychology*, *Australian Journal of Parapsychology*, *European Journal of Parapsychology* (including the *Research Letter of the Utrecht University Parapsychology Laboratory*), and the *Journal of Scientific Exploration*.

An exhaustive search was conducted of research databases including PsycINFO, Google Scholar, WorldCat, and LexScien, using the keywords "individual differences," "precognition," "parapsychology," "forced-choice," "retrocausation," "retrocausality," "psi," "ESP," and "extrasensory perception." Most of

these searches located studies that were already found in the journals listed above.

The search period was intended to capture all experimental psi research published from 1945 through 2016 that included individual difference measures.

The search strategy revealed 35 individual difference variables including more common measures such as extraversion and belief in psi, along with less widely used measures such as temporal lobe dysfunction and latent inhibition.

Selection Criteria

Studies were included from 1945 until 2016 if they met the following criteria:

1. Forced-choice design
2. Precognition design
3. Included individual difference measure(s)
4. A minimum of two human participants

Studies that did not include relevant information were excluded. For example, the results reported in Wiseman and Greening (2002) could not be included as their precognition and clairvoyance data were combined when reporting individual difference measures (e.g., the sheep-goat effect), and results reported by Steinkamp (1998) could not be used as the number of participants was not reported.

The identification, screening, and eligibility of the studies followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al., 2009). Figure 1 provides a detailed summary of the database search and screening process.

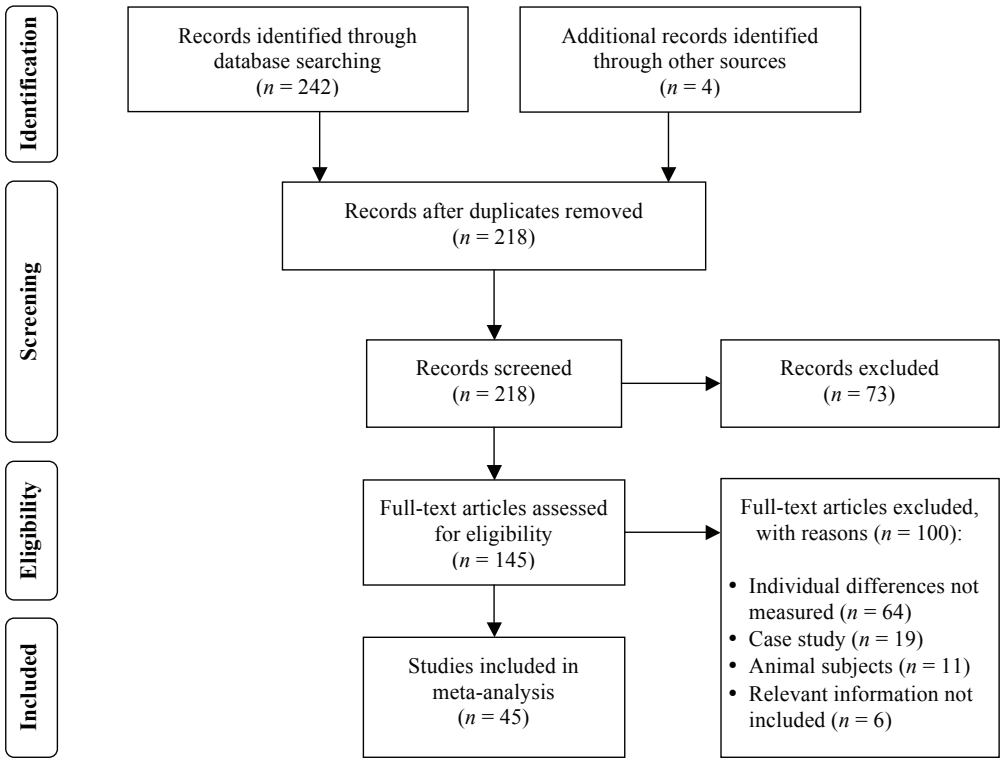


Figure 1. A flow diagram illustrating the database search and screening stages involved in this meta-analysis.[2]

[2] In both the screening stage and the eligibility stage, a number of articles (73 and 64, respectively) were excluded as they did not measure individual differences and were thus not relevant to this meta-analysis. The only difference is that the articles identified for exclusion in the screening stage were deemed irrelevant after only screening the titles, abstracts, and keywords, whereas the others required a more thorough reading of the article to make that determination.

Definitions

Independent investigator. For the purposes of this meta-analysis, an "investigator" refers to the lead author of a study. An investigator is considered independent of another investigator if both investigators have never worked on a paper together using the database being analysed (and/or worked with the other investigator's co-authors). The number of independent investigators can be helpful in determining how replicable an effect might be; the fewer independent investigators there are—even if the studies have been repeated multiple times—the less certain we can be that the results are replicable. As Hyman (1977) notes, it is not enough to simply repeat the same results, for whatever errors or biases may have occurred in the first instance might also be part of the subsequent repetitions of the experiment.

Individual differences. For the purposes of this research, an individual difference is defined as anything that an individual may psychologically vary on, whether it is personality, beliefs, intelligence, or aptitude (Nazimuddin, 2015). However, this meta-analysis makes a distinction between individual differences that are relatively constant regardless of situational factors (e.g., trait-level individual differences) and those that are more temporary and can be affected by the experimental situation (e.g., state-level individual differences such as participants' mood in the experiment); the latter (e.g., classification of a participant high in trait-anxiety but low in state-anxiety in a particular experimental setting) will not be included in this paper to reduce confusion while also limiting the number of variables analysed.

Individual differences were further categorised, combining similar measures (or subcomponents) into families of similar constructs. Where multiple measures of a single individual difference were used in a study, only the most appropriate measure was used. For example, some studies included both a sheep-goat measure and an interest in psi measure—in these cases, only the sheep-goat measure was included, as it is has historically been more consistently used as a measure of psi belief (Lawrence, 1993).

Procedural Features

As there are a multitude of individual difference measures included (35 individual difference constructs were identified for analysis), a separate meta-analysis was performed on each category of individual difference. The majority of these meta-analyses contain less than five studies in total, so it was not practical to code for procedural features.

Quality coding of the studies was not implemented for four reasons. Firstly, up until 1976, the founder of experimental parapsychology, J. B. Rhine, encouraged less detail in publications for nonsignificant parapsychological findings than significant findings (Steinkamp, 2005). Therefore, quality coding would inevitably favour newer studies, since more information is available for post-1976 studies (which may not correlate with the actual quality of the experiment). Secondly, precognition experiments have less potential for procedural defects compared with parapsychological research more generally, which is reflected by Honorton and Ferrari (1989) having only 6 quality criteria for precognition experiments rather than Rhine, Pratt, Stuart, Smith, and Greenwood's (1940) 34 quality criteria for ESP experiments. Thirdly, Honorton and Ferrari (1989) did not find a relationship between forced-choice precognition experiments and their quality ratings. Lastly, the majority of these meta-analyses had too few studies in total to meaningfully differentiate them on quality.

However, year of publication was coded, as it allowed us to examine whether effect sizes have increased over time, stayed the same, or even decreased. Honorton and Ferrari (1989) suggest that if effect sizes do not increase over time—as they found in their meta-analysis—it might mean that researchers lack an understanding of the underlying factors of psi performance (since they could not reliably increase its magnitude over time). Alternatively, an increase in effect size over time would be more promising, as was reported in Storm et al.'s (2012) meta-analysis.

Meta-Analysis of Correlation Coefficients

All indices of association between an individual difference measure and psi performance were converted to correlation coefficients using Comprehensive Meta-Analysis (CMA) software version 2 (Boren-

stein, Hedges, Higgins, & Rothstein, 2005) or manually. For example, *t* tests were converted to point-biserial correlations and phi coefficients were computed from 2 x 2 contingency tables. Some studies gave only trial-based data such as the critical ratio (*z*; e.g, Buzby 1967; Freeman & Nielsen, 1964). In these instances, correlations were estimated using a method for estimating effect sizes from critical ratios described by McCarthy and Schechter (1986), providing an estimate of Cohen's *d*—this was then converted to the *r* metric. Unreported correlations were estimated using the provided *p* values, whereas studies that reported only nonsignificance had their correlation set to .00,[3] a practice consistent with the approach adopted by Honorton et al. (1998) in their meta-analysis of extraversion and ESP performance. Where necessary, correlation signs were adjusted to reflect the appropriate relationship between the individual difference measure and psi performance. Finally, CMA weighted each study—using a random effects model incorporating both sample size and between-study variance—giving an overall outcome metric (*r*) for each individual difference measure in the database. A random effects model was used rather than a fixed effects model, as most studies were not exact replications of each other and this model takes into account such variation (Borenstein, Hedges, Higgins, & Rothstein, 2010). All *p* values are two-tailed.

Heterogeneity tests using Cochran's *Q* were also conducted on each meta-analysis to determine whether results from the included studies were representative of a single homogenous effect.[4] For those meta-analyses showing heterogeneity, moderator analyses were conducted using the year of publication as a proxy for methodological quality. The I^2 index was also reported, to give an idea of the degree of heterogeneity present. Finally, Rosenthal's (1979) fail safe *N*, or the file drawer estimate, was calculated for all meta-analyses that showed statistical significance to determine how many unreported studies averaging null results would need to exist for the effect to be reduced to overall nonsignificance. If the number is high, then there is less likelihood for publication bias, that is, studies being reported only if they show statistical significance (Honorton & Ferrari, 1989). Because unreported nonsignificant studies may have an average effect size below zero (Ferguson & Heene, 2012), an alternative method for examining publication bias, namely Egger's regression method, was also included for these studies (Egger, Smith, Schneider, & Minder, 1997). Egger's test aims to quantify potential asymmetrical distributions of studies around the mean effect size (Rothstein, 2008).

Results

Descriptives

Overall, this meta-analysis is comprised of 55 individual studies, which were reported in 45 papers and conducted by 17 independent investigators. The studies span a total of 71 years, between 1945 and 2016. There were a total of 17,584 participants analysed, with sample sizes ranging from 13 to 13,941. In the majority of these studies, students were the sample population.

Separate meta-analyses of the relationship between psi performance and each category of individual differences are reported below. It is ordered in terms of those variables that have the most exemplar studies (from belief in ESP, the Big Five, various operationalisations of luck, which have the most studies) through to variables for which there are only two or three studies (e.g., religiosity, emotional reactivity, intelligence).

For a summary of the total number of studies, independent investigators, and participants, see Table 2.

Major Individual Difference Measures

[3] However, this is an estimate, as the mean of the distribution of all possible nonsignificant outcomes is likely to be less than zero after removing the outcomes that give significant results by chance (as it effectively removes or truncates the right tail of the distribution). See the Appendix, Table A1, for all studies that this applies to.

[4] Note that these tests were conducted for all meta-analyses, even when there were only two or three studies, as it would be hard to justify an arbitrary cut-off point. However, it does not imply that all of these tests should be given equal weight. The heterogeneity analyses conducted on a limited number of studies should not be considered definitive.

Belief in ESP. Overall, belief in ESP was the most studied potential individual difference correlate in forced-choice precognition experiments, having been reported in 22 studies by 12 independent investigators based on a total of 2,200 participants. The most common measurement questionnaire was a variant of Schmeidler's (1943) sheep-goat criterion, such as Thalbourne and Delin's (1993) *Australian Sheep-Goat Scale* or Bahdra's (1966) *Sheep-Goat Questionnaire*. In general, participants who score high on these scales are classified as "sheep" (or believers in ESP) and those who score low are classified as "goats" (or disbelievers in ESP).

Figure 2 shows a forest plot of the correlation coefficients, with the correlations ranging from -.17 to .72. The overall mean weighted effect size (r) is .13 (p = .002), with a 95% confidence interval between .05 and .20. This suggests that there is a small but significant relationship between psi belief and performance on a psi task, such that people who believe in psi tend to perform better than those who do not believe in psi. This effect size is slightly larger than the effect size reported by Lawrence (1993) in his meta-analysis on the sheep-goat effect (r = .03), but that also included telepathy and clairvoyance experiments.

Study name	Statistics for each study			
	Correlation	Lower limit	Upper limit	Total
Schmeidler, 1945	0.04	-0.19	0.26	75
Buzby, 1967	0.30	0.17	0.42	202
Ryzl, 1968a	0.29	0.04	0.50	63
Ryzl, 1968b	0.31	-0.09	0.63	25
Johnson, 1969	0.62	0.33	0.81	28
Nielsen, 1970	0.72	0.28	0.91	13
Honorton, 1972	0.00	-0.37	0.37	28
Haraldsson, 1975	0.00	-0.13	0.13	223
Haraldsson, 1980	0.11	0.02	0.20	449
Thalbourne et al., 1982 (Exp 1)	0.00	-0.20	0.20	101
Thalbourne et al., 1982 (Exp 2)	0.30	0.09	0.48	86
Thalbourne, 1996	-0.06	-0.25	0.14	99
Haraldsson et al., 2002	0.06	-0.22	0.33	50
Storm, 2006	0.10	-0.04	0.24	199
Wilson & Hanlin, 2007 (Exp 1)	0.02	-0.24	0.27	61
Wilson & Hanlin, 2007 (Exp 2)	0.00	-0.30	0.30	44
Storm, 2008	-0.17	-0.32	-0.01	149
Luke et al., 2008	0.24	0.04	0.41	100
Luke & Morin, 2009	0.49	0.21	0.69	41
Palmer, 2009	0.00	-0.25	0.25	64
Hitchman et al., 2012	-0.03	-0.31	0.25	50
Hitchman et al., 2015 (Combined)	0.01	-0.27	0.28	50
	0.13	0.05	0.20	2200

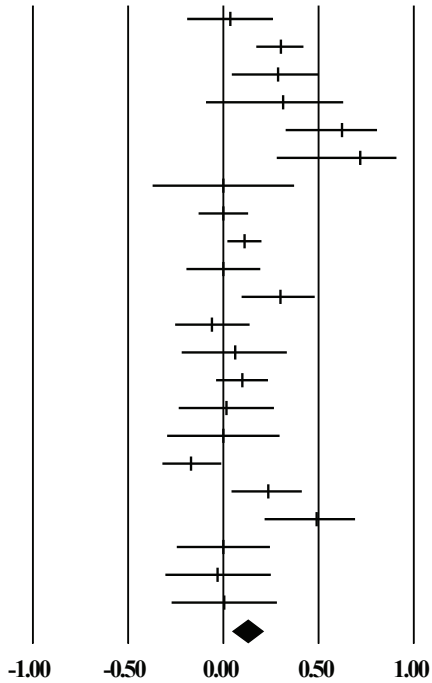


Figure 2. Meta-analysis of the relationship between belief in ESP and psi performance in forced-choice precognition experiments.

However, a test of heterogeneity was significant (Q = 61.16, p < .001) which suggests that variation in results may be due to factors other than the relationship between psi belief and performance (for example, error, or the influence of a moderator). The I^2 was 66%. Consequently, a mixed effects model (method of moments) meta-regression was conducted, which found year of publication to be a significant moderator, QR = 6.71, p = .01. Figure 3 shows effect sizes to decrease as year of publication increases. This means that some of the heterogeneity that caused the significant Q can be attributed to the year of publication.

Finally, the fail safe N, or the number of unreported studies averaging null results that would be needed to bring the p value to nonsignificance, is 141. Egger's test was not found to be significant, $t(20)$ = 1.21, p = .24.

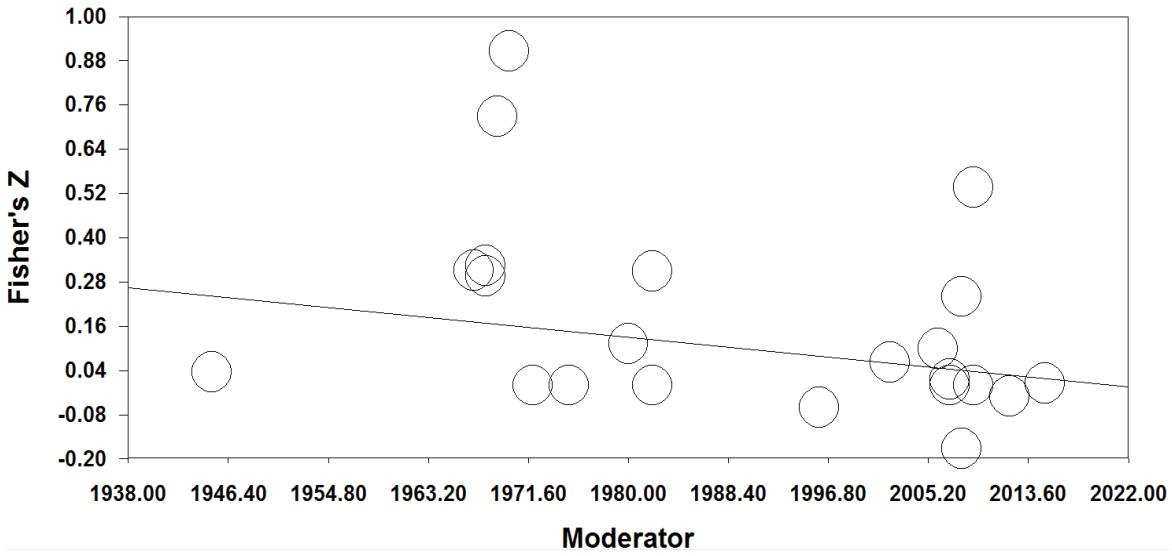


Figure 3. Meta-regression on the relationship between belief in ESP and psi performance in forced-choice precognition experiments, using publication date as the moderator.

The Big Five

The next set of meta-analyses cover personality indicators consistent with the Big Five personality traits (McCrae & Costa, 1987): extraversion (how outgoing and social a person is), neuroticism (a long-term tendency to be in a negative emotional state, such as anxious or frustrated), openness to experience (intellectually curious, willing to try new things, and imaginative), agreeableness (how kind and sympathetic a person is), and conscientiousness (organised and diligent).

Extraversion. Extraversion was the second most studied individual difference measure in forced-choice precognition experiments, having been reported in 14 studies by seven independent investigators with a total of 1,206 participants. Extraversion was typically measured within larger personality questionnaires such as the *16PF* (Cattell & Mead, 2008), but subcomponents such as Bem's (2011) *Sensation Seeking Scale* were also included. High scorers on these measures are generally considered to be extraverted and low scorers introverted.

Figure 4 shows a forest plot of the correlation coefficients, with the correlations ranging from -.28 to .35. The overall mean weighted effect size (r) is .08 (p = .02), with a 95% confidence interval between .01 and .15. This suggests that there is a small but significant relationship between extraversion and psi performance, such that people who are extraverted tend to perform better than those who are more introverted. This result is consistent with previous studies that have also found a positive relationship between extraversion and psi performance (Mangan, 1958; Palmer, 1978; Honorton et al., 1998). Furthermore, a test of heterogeneity was not significant (Q = 17.23, p = .19), with an I^2 index of 25%.

Finally, the fail safe N, or the number of unreported studies averaging null results that would be needed to bring the p value to nonsignificance, is nine, whereas Egger's test was not significant , $t(12)$ = 0.56, p = .59.

Neuroticism. Neuroticism was measured using a variety of different questionnaires encompassing anxiety, affect, and mood, and was included in nine studies by seven independent investigators and a total of 528 study participants. The correlation coefficients range from -.38 (Humphrey, 1945) to .60 (Freeman & Nielsen, 1964). The overall mean weighted effect size (r) is .05 (p = .43), with a 95% confidence interval between -.08 and .19. The results are inconclusive about whether an actual effect occurs, falling slightly short of Steinkamp's (2005) suggestion that neuroticism was a promising predictor of ESP forced-choice experiments.

Study name	Correlation	Lower limit	Upper limit	Total
Humphrey, 1945	0.35	-0.12	0.69	19
Nielsen, 1970	0.00	-0.55	0.55	13
Thalbourne et al., 1982 (Exp I)	0.00	-0.20	0.20	101
Thalbourne et al., 1982 (Exp II)	0.00	-0.21	0.21	86
Storm & Thalbourne, 1989-99	0.28	0.08	0.46	93
Storm & Thalbourne, 2001	0.00	-0.19	0.19	107
Haraldsson et al., 2002	-0.03	-0.31	0.25	50
Storm, 2002	-0.03	-0.33	0.27	43
Wilson & Hamlin, 2007	-0.28	-0.53	0.02	44
Luke et al., 2008	0.16	-0.03	0.35	100
Bem, 2011 (Exp 1)	0.18	-0.02	0.36	100
Bem, 2011 (Exp 2)	0.17	0.01	0.32	150
Wagenmakers et al., 2012	0.13	-0.07	0.32	100
Thalbourne & Storm, 2014	0.04	-0.10	0.18	200
	0.08	0.01	0.15	1206

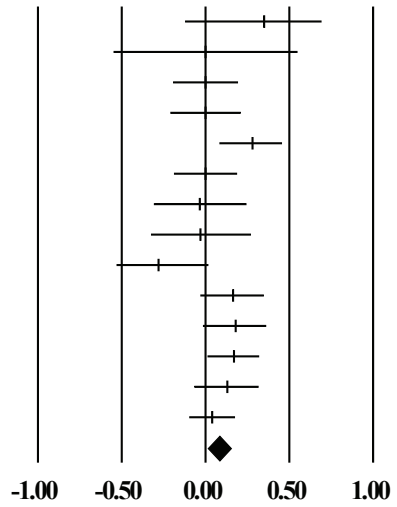


Figure 4. Meta-analysis of the relationship between extraversion and psi performance in forced-choice precognition experiments.

However, a test of heterogeneity was significant (Q = 17.29, p = .03) which suggests that there were potential moderating factors in this database. The I^2 was 54%. A mixed effects model (method of moments) meta-regression did not find year of publication to be a significant moderator, QR = 0.30, p = .58. Due to meta-regression analyses not being recommended for meta-analyses with less than 10 studies (Borenstein, Hedges, Higgins, & Rothstein, 2009), this finding should be treated with caution.

Openness to experience. Openness to experience was reported in nine studies of 522 participants, by five independent investigators. The most common measurement questionnaire was the *Openness to Experience Scale* (Goldberg, 1999). Figure 5 shows a forest plot of the correlation coefficients, with the correlations ranging from -.08 to .46. The overall mean weighted effect size (r) is .12 (p = .02), with a 95% confidence interval between .02 and .22, indicating a small but significant relationship between openness to experience and psi performance, such that people who prefer new experiences tend to perform better than those who prefer familiar routines. Furthermore, a test of heterogeneity was not significant (Q = 11.56, p = .24), with an I^2 index of 22%.

Study name	Correlation	Lower limit	Upper limit	Total
Storm & Thalbourne, 1998-99	0.00	-0.20	0.20	93
Haraldsson et al., 2002	0.16	-0.12	0.42	50
Wilson & Hamlin, 2007	0.18	-0.12	0.45	44
Luke et al., 2008	0.46	0.13	0.70	32
Luke & Morin, 2009	-0.08	-0.38	0.23	41
Palmer, 2009	0.27	0.02	0.49	61
Hitchman et al., 2012	0.29	0.01	0.53	50
Hitchman et al., 2015a	0.01	-0.27	0.29	49
Hitchman et al., 2015b (Combined)	-0.01	-0.29	0.27	50
Hitchman et al., 2016	0.06	-0.22	0.33	52
	0.12	0.02	0.22	522

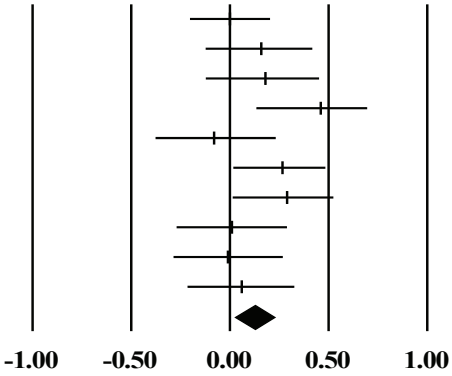


Figure 5. Meta-analysis of the relationship between openness to experience and psi performance in forced-choice precognition experiments.

In this case, the fail safe *N*, or the number of unreported studies averaging null results that would be needed to bring the *p* value to nonsignificance, is 12. Egger's test was not significant, $t(8) = 1.30$, $p = .23$.

Agreeableness. Agreeableness was reported in seven studies by four independent investigators with a total of 556 participants. The most common measurement questionnaire was the Independence factor of the *16PF* (Cattell, 1996; Cattell & Mead, 2008). Correlation coefficients range from -.36 (Humphrey, 1945) to .23 (Storm & Thalbourne, 1998-1999), with an overall mean weighted effect size (*r*) of .02 ($p = .71$) and a 95% confidence interval between -.09 and .13. Although the results are inconclusive, the data suggest that the true effect size is below .13 and could be zero. There was no indication of significant heterogeneity ($Q = 9.34$, $p = .16$), with an I^2 index of 36%.

Conscientiousness. Conscientiousness was the least studied Big Five personality trait in forced-choice precognition experiments, having been reported in only three studies by three independent investigators with a total of 187 participants. Conscientiousness was measured as a component of larger personality questionnaires such as the *NEO Five-Factor Inventory* (McCrae & Costa, 1987). Correlations range from .00 (Storm & Thalbourne, 1998-1999) to .23 (Wilson & Hamlin, 2007). The overall mean weighted effect size (*r*) is .06 ($p = .45$), with a 95% confidence interval between -.09 and .20, which is inconclusive but suggests that the effect size may lie below .20 and could be zero. Furthermore, a test of heterogeneity was not significant ($Q = 1.69$, $p = .43$). The I^2 is 0%.

Beliefs About Luck

Although luck has been explored in multiple studies, Smith (1998) showed that "luck" can mean different things to different people. Therefore, various measurements of luck and luck beliefs have been reported in forced-choice precognition experiments that include not only perceived luckiness, but also controllable luck belief, chance belief, providence belief, and fortune belief. The most common tool of measurement used in these experiments was Luke, Delanoy, and Sherwood's (2003) *Questionnaire of Beliefs about Luck*, which incorporates all of these subcomponents together. They will now be discussed in turn.

Perceived luckiness. Perceived luckiness has (prior to 2008) been the standard measurement used to explore luck in psi experiments and refers to how lucky one perceives oneself to be. For forced-choice precognition experiments, perceived luckiness was reported in four studies by two independent investigators with a total of 231 participants. Correlations range from -.20 (Hitchman, Row, & Sherwood, 2012) to .26 (Luke, Delanoy, & Sherwood, 2008), with an overall mean weighted effect size (*r*) of .08 ($p = .49$); 95% CI [-.14, .28]. These results are inconclusive but suggest that the effect size is below .28 (and could be zero). A test of heterogeneity was not significant ($Q = 7.09$, $p = .07$), with an I^2 index of 58%.

Luck belief. Luck belief refers to the belief that luck is primarily controllable, and participants who score high in this belief also view luck as internal, stable, and nonrandom (Luke et al., 2003). Luck belief was reported in five studies by one independent investigator with a total of 248 participants. Figure 6 shows a forest plot of the correlation coefficients, ranging from -.09 to .26. The overall mean weighted effect size (*r*) is .13 ($p = .048$), with a 95% confidence interval between .001 and .26, indicating a small but reliable relationship between luck belief and psi performance, such that people who believe luck to be controllable tend to perform better than those who see luck as uncontrollable. Furthermore, a test of heterogeneity was not significant ($Q = 4.11$, $p = .39$). The I^2 is 3%. Finally, the fail safe *N*, or the number of unreported studies averaging null results that would be needed to bring the *p* value to nonsignificance, is less than 1. However, Egger's test is not significant, $t(3) = 1.34$, $p = .27$.

Chance belief. Chance belief refers to the belief that luck is random, unpredictable, unstable, and inert (Luke et al., 2003). Chance belief was reported in five studies by one independent investigator with a total of 248 participants. Correlations range from -.16 (Luke et al., 2008) to .48 (Luke, Roe, & Davison, 2008). The overall mean weighted effect size (*r*) is .14 ($p = .23$), with a 95% confidence interval between -.09 and .36.

Study name	Statistics for each study			
	Correlation	Lower limit	Upper limit	Total
Luke et al., 2008a	0.26	0.07	0.44	100
Luke et al., 2008b (Study 1)	0.14	-0.27	0.51	25
Luke et al., 2008b (Study 2)	0.12	-0.24	0.45	32
Luke & Morin, 2009	-0.09	-0.39	0.22	41
Hitchman et al., 2012	0.04	-0.24	0.31	50
	0.13	0.00	0.26	248

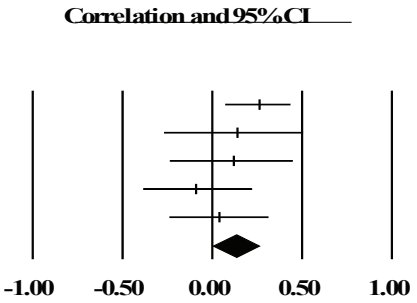


Figure 6. Meta-analysis of the relationship between luck belief and psi performance in forced-choice precognition experiments.

However, a test of heterogeneity was significant (Q = 11.40, p = .02) which suggests that there were potential moderating factors in this database. The I^2 is 65%. A mixed effects model (method of moments) meta-regression was conducted, but it did not find year of publication to be a significant moderator (QR = 1.75, p = .19).

Providence belief. Providence belief refers to the belief that luck is something that is managed by external forces or higher beings (Luke et al., 2003). Providence belief was reported in five studies by one independent investigator with a total of 248 participants. Correlations range from -.09 (Hitchman, Roe, & Sherwood, 2012) to .39 (Luke et al., 2008). The overall mean weighted effect size (r) is .12 (p = .11; 95% CI [-.03, .27]). Furthermore, a test of heterogeneity was not significant (Q = 5.34, p = .25). The I^2 is 25%.

Fortune belief. Fortune belief refers to the belief that luck is meant as a metaphor for life's successes rather than as a literal event (Luke et al., 2003) and was reported in five studies by one independent investigator based on 248 participants. Correlations range from -.08 (Luke & Morin, 2009) to .15 (Luke et al., 2008). The overall mean weighted effect size (r) is .03 (p = .62; 95% CI [-.10, .16]). A test of heterogeneity was not significant (Q = 1.98, p = .74), and the I^2 index is 0%.

Uncategorised Individual Difference Measures

Creativity. Creativity was reported in nine studies by three independent investigators with a total of 506 participants. The most common measurement questionnaires were the short version of the *Personal-Social Motivational Inventory* (Torrance, 1963) and the *Creative Cognition Inventory* (Holt, 2002). Correlations range from -.17 (Schmeidler, 1964c) to .20 (Luke et al., 2008), with an overall mean weighted effect size (r) of .05 (p = .46; 95% CI [-.08, .17]). Furthermore, a test of heterogeneity was not significant (Q = 5.90, p = .21), with an I^2 index of 32%.

Perceptual defensiveness. Perceptual defensiveness refers to psychological defence mechanisms and is related to subliminal perception and preconscious processing. All reported studies administered the *Defense Mechanism Test* (Kragh & Smith, 1970), with a total of six studies conducted by one independent investigator and a total of 272 participants. The test incorporates a tachistoscopic technique using peripheral stimuli to trigger subliminal anxiety and thereby defensive reactions. Figure 7 shows a forest plot of the correlation coefficients, with the correlations ranging from -.04 to .30. The overall mean weighted effect size (r) is .12 (p = .049; 95% CI [.001, .24]), suggesting a small but significant relationship between perceptual defensiveness and psi performance, such that people who exhibit high preconscious defensiveness tend to perform better than those who do not. There was no evidence of heterogeneity (Q = 5.13, p = .40), with an I^2 index of 3%.

Transliminality. Transliminality is defined as "the hypothesised tendency for psychological material to cross thresholds into or out of consciousness" (Thalbourne & Delin, 1994, p. 31), and was used in five studies by one independent investigator based on a total of 542 participants, with the most common measurement questionnaire being the *Transliminality Scale* (Thalbourne, 1998). Correlations range from

-.13 (Thalbourne, 1996) to .27 (Storm & Thalbourne, 1998-1999). The overall mean weighted effect size (*r*) is .01 ($p = .91$), with a 95% confidence interval between -.13 and .15. Although these results are inconclusive, a test of heterogeneity was significant ($Q = 9.81$, $p = .04$), but year of publication was not a significant moderator ($QR = 1.29$, $p = .26$). Due to the small number of studies, this finding should be treated with caution (Borenstein et. al., 2009). The I^2 is 59%.

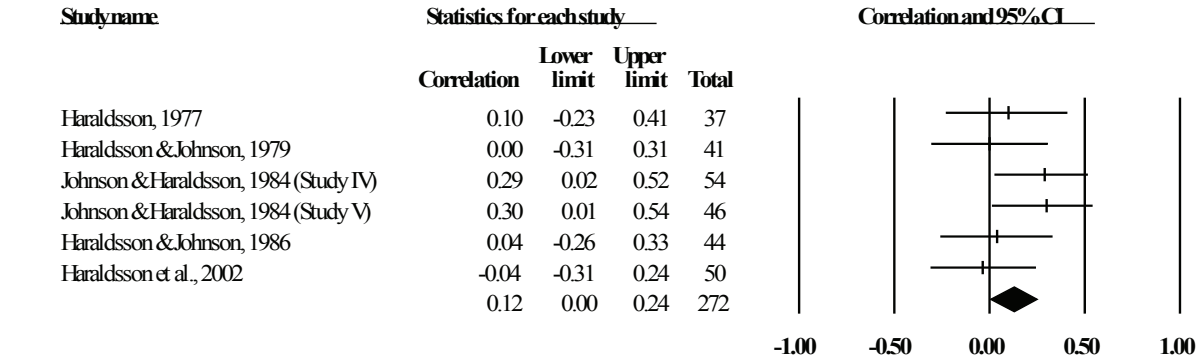


Figure 7. Meta-analysis of the relationship between perceptual defensiveness and psi performance in forced-choice precognition experiments.

Pro attitude. Pro attitude refers to "an attitude that is favourably directed towards an outcome" (Storm, 2002, p. 47) and three studies with 393 participants were conducted by a single investigator. All used Thalbourne and Storm's (2014) *Pro Attitude Scale*. Correlations range from -.16 (Storm, 2002) to .02 (Storm, 2008), with an overall mean weighted effect size (*r*) of -.02 ($p = .70$; CI 95% [-.12, .08]). A test of heterogeneity was not significant ($Q = 1.04$, $p = .60$), with an I^2 of 0%.

Dream recall. Dream recall (specifically, whether an individual recalls their dreams or not) was reported in four studies by three independent investigators with a total of 799 participants. It was typically measured using a one-item questionnaire, which asked participants how frequently they recalled their own dreams. Correlations range from .03 (Thalbourne, 1996) to .43 (Honorton, 1972). The overall mean weighted effect size (*r*) is .07 ($p = .23$; 95% CI [-.04, .18]), and a test of heterogeneity was not significant ($Q = 5.84$, $p = .12$). The I^2 is 49%.

Reports of unusual spontaneous experiences. Unusual spontaneous experiences can be described as seemingly paranormal experiences in everyday life (as opposed to the experimental laboratory) and were reported in three studies by two independent investigators based on a total of 695 participants. It was typically measured using a single question, which asked participants if they have had any precognitive dreams (i.e., dreams that they thought predicted the future). Correlations range from -.21 (Schmeidler, 1964d) to .00 (Haraldsson 1975, 1980), with an overall mean weighted effect size (*r*) of -.01 ($p = .87$; 95% CI [-.08, .07]). A test of heterogeneity was not significant ($Q = 0.85$, $p = .65$), with an I^2 of 0%.

Religiosity. Religiosity was reported in two studies by two independent investigators with a total of 149 participants. In both studies, religiosity was measured using the *Religiosity Scale* (Haraldsson, 1993), with the correlations ranging from -.13 (Thalbourne, 1996) to .08 (Haraldsson, Houtkooper, Schneider, & Bäckström, 2002). The overall mean weighted effect size (*r*) is -.05 ($p = .59$; 95% CI [-.24 and .14]). A test of heterogeneity was not significant ($Q = 1.34$, $p = .25$). The I^2 is 26%.

Emotional reactivity. Emotional reactivity is a measure of one's emotional reaction to violent, scary, or gruesome content in photographs, movies, and videos. All studies used the *Emotional Reactivity Scale* (Bem, 2003). A total of three studies by one investigator looked at emotional reactivity, with 151 participants being included in the experiments. Correlations range from -.27 (Hitchman, Sherwood, & Roe, 2015) to .29 (Hitchman, Pfeuffer, Roe, & Sherwood, 2016), with an overall mean weighted effect size (*r*) of .06 ($p = .71$; 95% CI [-.27, .38]). A test of heterogeneity was significant ($Q = 8.53$, $p = .01$), with the year of publication being a significant moderator ($QR = 3.96$, $p = .046$). Figure 8 shows effect sizes to increase as year of publication increases. The I^2 is 77%.

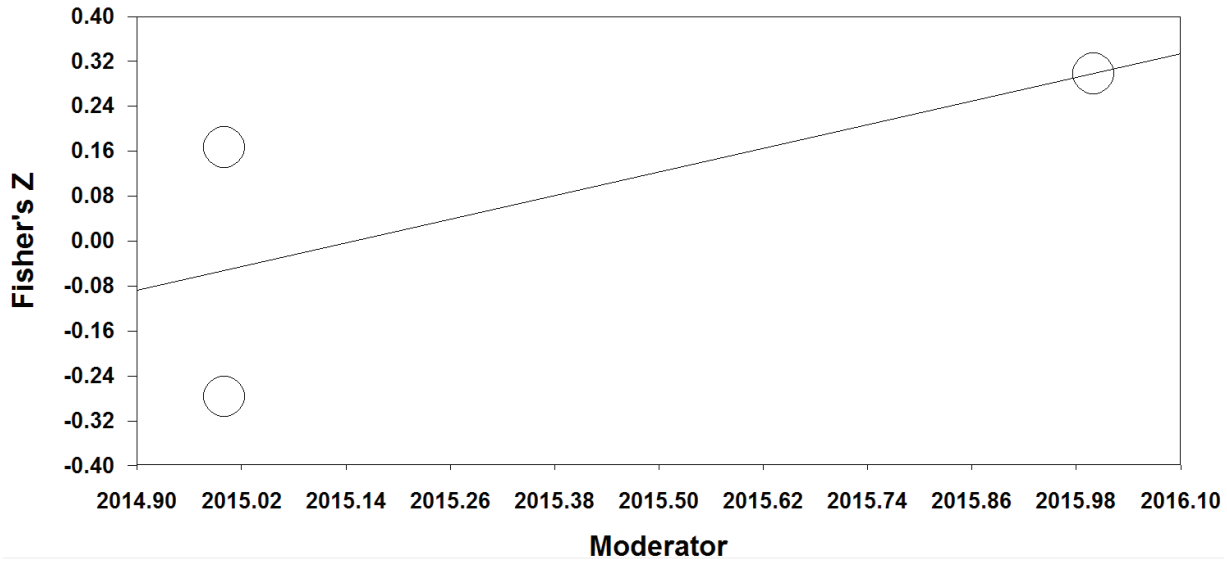


Figure 8. Meta-regression on the relationship between emotional reactivity and psi performance in forced-choice precognition experiments, using publication date as the moderator.

Temporal lobe dysfunction. Temporal lobe dysfunction measures symptoms of temporal lobe damage such as disturbances of perception, selective attention of auditory input, and impaired organisation of verbal material. A total of two studies by two independent investigators looked at temporal lobe dysfunction, using either the 13-item *LIMBEX Scale* or the Complex Partial Epileptic Signs cluster of the *Personal Philosophy Inventory* (Persinger & Makarec, 1987). There were a total of 114 participants in all of the experiments. Correlations range from -.01 (Hitchman, Roe, & Sherwood, 2015) to .00 (Palmer, 2009), with an overall mean weighted effect size (r) of -.004 (p = .96; 95% CI [-.19, .18]). A test of heterogeneity was not significant (Q = 0.003, p = .96), with an I^2 of 0%.

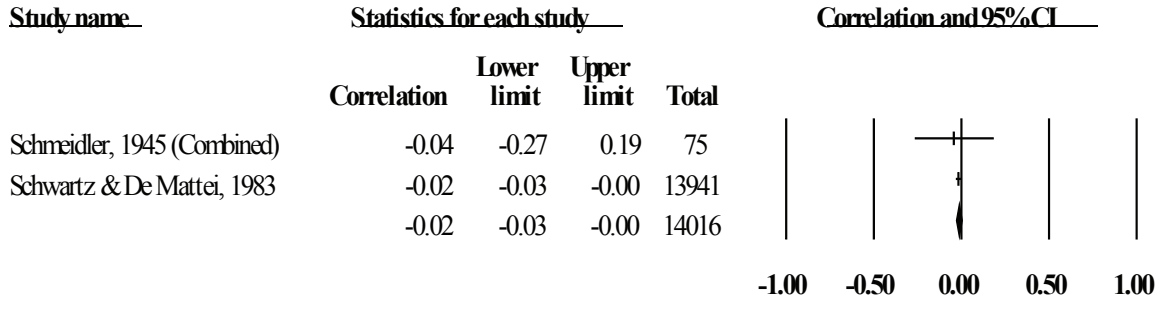


Figure 9. Meta-analysis of the relationship between time belief as dynamic and psi performance in forced-choice precognition experiments.

Time belief as dynamic. "Time belief as dynamic" was reported in two studies by two independent investigators with a total of 14,016 participants and refers to how strongly one sees time as being in a constant flow. In both studies, Knapp and Garbutt's (1958) *Time Metaphor Test* was administered, which measures the belief that time is metaphorically "a dashing waterfall" or "a fast moving shuttle," for example. Figure 9 shows a forest plot of the correlation coefficients, with the correlations ranging from -.04 to -.02. The overall mean weighted effect size (r) is -.02 (p = .04; (95% CI [-.03, -.001]), suggesting a small but significant negative relationship between time belief as dynamic and psi performance, such that people who do not view time as dynamic tend to perform better than those who see time as dynamic and hasty.

Furthermore, a test of heterogeneity was not significant ($Q = 0.05$, $p = .82$), and the I^2 is 0%. The fail safe N was not calculated, as it is impractical to do so with less than three studies.

Future-orientation. Future-orientation refers to being more attentive towards future events than past events and was reported in two studies by two independent investigators based on a total of 118 participants. Both studies used the *Attitude Toward the Future Questionnaire* (Vaughan & Houck, 1993). Correlations range from -.04 (Haraldsson et al., 2002) to .12 (Vaughan & Houck, 1993), and the overall mean weighted effect size (r) is .05 ($p = .57$; 95% CI [-.13, .23]). A test of heterogeneity was not significant ($Q = .65$, $p = .42$), with an I^2 of 0%.

Intelligence. Intelligence was measured in two studies by two independent investigators with a total of 80 participants. Both studies included only child participants, ranging in age from 4 to 14. The measurement tools used were the *Peabody Picture Vocabulary Test* (Dunn & Hottel, 1961) and a formal mathematical ability test using basic addition, subtraction, multiplication, and division problems. Correlations range from -.23 (Winkelman, 1981) to .04 (Drucker, Drewes, & Rubin, 1977), with an overall mean weighted effect size (r) of -.07 ($p = .62$; 95% CI [-.31, .19]). A test of heterogeneity was not significant ($Q = 1.2$, $p = .27$), and the I^2 is 19%.

Single Studies

There were also a number of individual difference measures that were analysed only for a single study. As single studies cannot be meta-analysed, they will be presented individually in Table 1 below, in order of effect size.

Table 1

Summary Statistics of All Individual Difference Measures Included in Only One Study

Individual Difference Measure	Study Author(s)	Effect Size (r)	Sample Size
Memory	Stanford, 1970	.36*	30
Sensitivity to Punishment	Hitchman et al., 2016	-.21	51
Optimism	Haraldsson et al., 2002	.20*	50
Fantasy Proneness	Thalbourne, 1996	-.13	99
Latent Inhibition	Hitchman , Sherwood, & Roe, 2015	-.11	48
Belief in the Occult	Haraldsson et al., 2002	.10	50
Time Belief as Naturalistic	Schmeidler, 1964b	-.08	75
Psychotism	Haraldsson et al., 2002	-.08	50
Hypersensitivity	Thalbourne, 1996	.04	99
Time Belief as Humanistic	Schmeidler, 1964b	.01	75
Cerebral Lateralization	Palmer, 2009	.00	64
Sensitivity to Reward	Hitchman et al., 2016	.00	51

*p <.05

Summary of Results

Below (Table 2) is a summary of all of the individual difference measures that were meta-analysed for comparison. The individual difference measures are ordered by overall effect size.

Column 1 is the individual difference measure; Column 2 is the 95% confidence interval,[5] Column 3 is the overall effect size; Column 4 is the number of individual studies included in the meta-analysis; Column 5 is the number of independent investigators in the meta-analysis; Column 6 is total number of participants for all of the studies in the meta-analysis combined.

Table 2
Summary Statistics of All Individual Difference Measures That Were Meta-Analysed

Individual Difference Measure	95% Confidence Intervals	Overall Effect Size (*r*)	Number of Studies	Number of Independent Investigators	Total Number of Participants
Chance Belief	-.09, .36	.141	5	1	248
Luck Belief	.001, .26	.131*	5	1	248
Belief in Psi	.05, .20	.125*	22	12	2,200
Perceptual Defensiveness	.001, .24	.125*	6	1	272
Providence Belief	-.03, .27	.125	5	1	248
Openness to Experience	.02, .22	.124*	9	5	522
Extraversion	.01, .15	.080*	14	7	1,206
Perceived Luckiness	-.14, .28	.076	4	2	231
Dream Recall	-.04, .18	.070	4	3	799
Intelligence	-.31, .19	-.065	2	2	80
Emotional Reactivity	-.27, .38	.064	3	1	151
Conscientiousness	-.09, .20	.056	3	3	187
Neuroticism	-.08, .19	.054	9	7	528
Religiosity	-.24, .14	-.054	2	2	149
Future-Oriented	-.13, .23	.053	2	2	118
Creativity	-.08, .17	.047	9	3	506
Fortune Belief	-.10, .16	.032	5	1	248
Agreeableness	-.09, .13	.021	7	4	556
Pro Attitude	-.12, .08	-.019	3	1	393
Time Belief as Dynamic	-.03, -.001	-.017*	2	2	14,016
Transliminality	-.13, .15	.008	5	1	542
Spontaneous Experences	-.08, .07	-.006	3	2	695
Temporal Lobe Dysfunction	-.19, .18	-.004	2	2	114

*$p < .05$

Discussion

As we can see from the summary of results, the majority of these individual difference measures have not been extensively investigated, with the exception of belief in psi, extraversion, and neuroticism. It

[5] As effect size estimates based on previous research are inherently uncertain, confirmatory studies based on lower confidence intervals are less likely to overestimate the true effect size (Kennedy, 2016). Therefore 80% and 68% confidence intervals are provided in the Appendix, Table A2.

might therefore be argued that such a meta-analysis is unnecessary. However, without a meta-analysis, researchers will likely impose their own synthesis of the data, and a meta-analysis can provide greater clarity in this regard—even if it only incorporates two or three studies, as Valentine, Pigott, and Rothstein (2010) argue that all other synthesis techniques are less transparent and/or less likely to be valid. At the same time, it is not intended to stop researchers from exploring individual difference measures that may not yet seem promising, especially those that have been tested only a handful of times; it is merely given as a benchmark of past results.

With that being said, the results suggest that there may be only a small pool of individual difference measures that are robustly correlated with performance on a forced-choice precognition task. This is also consistent with Steinkamp's (2005) review of forced-choice ESP experiments, where she found that "there are few variables which have correlated clearly with success . . . most variables tested provided little evidence either way as being ultimately psi-conducive and there were relatively few variables that appeared to be encouraging" (p. 155). However, notable exceptions in this meta-analysis include extraversion and belief in psi, which show more consistent results across a larger number of studies. It should also be noted that with the number of meta-analyses conducted in this paper, there is an increased risk of family-wise error, and that one or more of these significant findings might be the result of multiple analyses (e.g., represent a false positive). Further, while forced-choice ESP tests produce normal distributions (unless the number of trials is very small and/or the number of response alternatives is very large), it may be useful to incorporate nonparametric statistics in the case of any highly skewed distributions. Researchers should bear this in mind when setting their expectations for future experiments.

Nevertheless, the findings suggest a small but significant relationship between the following individual difference measures and psi performance: luck belief (specifically, the belief that luck is primarily controllable), perceptual defensiveness, openness to experience, belief in psi, extraversion, and time belief as dynamic. Perhaps what underlies these individual differences is a mechanism derived from being open-minded, curious, social, and intuitive—all of which might lead people to discuss, think about, and explore the "paranormal." Consequently, these same people may act on information or intuitions that others may ignore or supress, leading them to make better decisions about the future than we would expect by chance alone. This may mean that a relationship exists between these variables and performance on a psi task, where such traits may either facilitate or innately allow demonstrations of psi.

However, given that even the strongest predictor (extraversion) in this meta-analysis accounts only for approximately 2% of the variance on its own, perhaps these predictors are not related and are instead additive, and provide more power when analysed together. Therefore, the best strategy for future researchers may be to combine individual difference factors, not just for the additive benefits but also to examine potential interactions (see Baron & Kenny, 1986) between the factors that may predict precognitive performance.

Alternatively, these results may be due to statistical anomalies, having arisen from the large amount of analyses being conducted on individual differences in psi research, or even due to methodological flaws. The other possibility is that the results reflect an actual relationship between certain variables combined with imperfect research designs. Taking into consideration the fact that many of these findings, including the nonsignificant results, were based on only a limited amount of studies, it is difficult to come to any strong conclusions. If one also considers the possibility of experimenter psi (i.e., where the experimenter influences the final results of an experiment due to his or her own psi abilities), it becomes extremely difficult to disentangle the data, especially in the case of a meta-analysis with only a single independent investigator. This potential explanation has previously been offered for the relationship between perceptual defensiveness and psi performance (see Haraldsson et al., 2002).

Retrospective meta-analyses also have several limitations, so it is not appropriate to make any definitive statements about the results without first conducting confirmatory studies. One such limitation of retrospective meta-analysis is that included studies are often affected by publication bias or the file-drawer effect, whereby only significant results are reported or published. Although parapsychological journals generally publish more nonsignificant results than most mainstream scientific journals (Mousseau, 2003), no field is entirely immune, especially when there may be tens or even hundreds of secondary analyses con-

ducted (e.g., various individual difference measures). Indeed, the low fail safe *N* numbers found in several of these meta-analyses (e.g., nine for extraversion) suggest that publication bias is a possibility. At the same time, there is no indication to argue strongly that publication bias is a problem when taking into account Egger's test results, which should be reassuring for parapsychologists given Mousseau's (2003) findings. Secondly, there will always be subjectivity involved in meta-analytical procedures and interpretation, such as defining and judging exclusion criteria, using search strategies, or coding the studies (Murray, 2011). Biases will come into play—whether conscious or unconscious—that influence procedural decisions, especially since researchers will already be aware of the results of the individual studies (Watt & Kennedy, 2016). This subjectivity allows psi proponents and critics to conduct meta-analyses whose conclusions often support their own prior beliefs, but never manage to convince the other side (Palmer, 2003).

Yet meta-analyses are still useful in that they can suggest the conditions under which replication is most likely to occur, assuming an effect exists at all. An overall effect size also gives future researchers the ability to calculate how many participants they would need to include in their experiment for it to be adequately powered. With these key pieces of information, prospective meta-analyses (which define the exclusion criteria and other details beforehand) can then be conducted using only future studies that are to be preregistered. A prospective meta-analysis therefore avoids all the potential issues of publication bias and subjectivity that are evident in a retrospective meta-analysis, while also addressing methodological issues such as optional stopping. In fact, Watt (2016) has recently set up a registration-based prospective meta-analysis of one of the most thorough yet controversial paradigms in parapsychology, the ganzfeld—Watt's (2016) meta-analysis protocol specifically includes only preregistered individual studies that prospectively fit their criteria. This is a positive direction for parapsychology, as it brings structure and focus to the field.

Preregistration has been made even easier by the Koestler Parapsychology Unit (KPU) registry, an initiative started at the University of Edinburgh in 2012 that allows researchers to prospectively register their experiments in detail, publically, and is not affiliated with a specific journal (Watt & Kennedy, 2015). Not only are prospective meta-analyses the ideal way to test the replicability of psi phenomena, but they are also the best way to confirm the null hypothesis should psi not exist. Alcock (2003) claims that the latter hypothesis often does not get serious consideration by parapsychologists, so multiple prospective meta-analyses showing nonsignificance may force parapsychologists to give the null hypothesis more deliberation than a single study or retrospective meta-analysis would.

Another consideration for attempting replication is the researcher conducting the experiment. Although some parapsychologists believe that the psi experimenter effect eliminates the possibility of true replication, that is, that due to the nature of psi only experimenters who are proponents of psi will get positive results in psi experiments whereas sceptics will not (Utts, 2015), most researchers would only be satisfied that psi phenomena exist if it were to be consistently demonstrated by neutral scientists and not just a select few who believe in psi (Alcock, 2003; Palmer, 2016). The current meta-analysis was conducted with this goal in mind, as forced-choice precognition experiments are arguably the easiest to run and can be automated using computer programmes. For example, Bem (2011) ran his Precognitive Detection of Erotic Stimuli experiment using an automated computer programme. This allows researchers to collect large amounts of data with relatively little effort, an important consideration if researchers are to try and replicate the small effect sizes shown in this meta-analysis (Steinkamp, 2005).

Ultimately, it is hoped that this meta-analysis can be used as a springboard for future research, allowing the findings to be used in a productive way and perhaps aiding in the development of research programmes that are specific and structured. As Watt (2005) comments, "Parapsychologists need to be far more systematic in how they tackle these questions. . . . Systematic follow-up is an essential prerequisite for demonstrating a replicable effect" (p. 222). With parapsychology being such a small field, it is important that researchers work together to build up a body of evidence that is considered respectable by both parapsychologists and mainstream academics. With the recent failures to replicate many foundational studies in both psychology and medicine (Open Science Collaboration, 2015), now is the perfect time to define what a replicable psi experiment really is and take advantage of the benefits of preregistration. Only then will

we be able to finally confirm or disconfirm some of the major hypotheses in psi research. Depending on whether you are extraverted or believe in psi, you may already know how it will turn out.

References

[*Paper in the meta-analysis]

Alcock, J. (2003). Give the null hypothesis a chance: Reasons to remain doubtful about the existence of psi. *Journal of Consciousness Studies*, *10*, 29–50.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

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. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi:10.1037/a0021524

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*, 4–18. doi:10.1037/0033-2909.115.1.4

Bhadra, B. H. (1966). The relationship of test scores in ESP. *Journal of Parapsychology*, *30*, 1–17.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis (version 2) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. Chichester, UK: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. doi:10.1002/jrsm.12

*Buzby, D. E. (1967). Subject attitude and score variance in ESP tests. *Journal of Parapsychology*, *31*, 43–50.*

Cattell, H. E. P. (1996). The original big-five: A historical perspective. *European Review of Psychology, 46*, 5–14.

Cattell, H. E., & Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (Vol. 2), (pp.135–178). Trowbridge, UK: Cromwell Press.

*Drucker, S. A., Drewes, A. A., & Rubin, L. (1977). ESP in relation to cognitive development and IQ in young children. *Journal of the American Society for Psychical Research, 71*, 289–298.

Dunn, L. M., & Hottel, J. V. (1961). Peabody Picture Vocabulary Test performance of trainable mentally retarded children. *American Journal of Mental Deficiency, 65*, 448–452.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629–634. doi: 10.1136/bmj.315.7109.629

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*, 555–561. doi:10.1177/1745691612459059

*Freeman, J., & Nielsen, W. (1964). Precognition score deviations as related to anxiety levels. *Journal of Parapsychology, 28*, 239–249.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*, 7–28.

*Haraldsson, E. (1975). Reported dream recall, precognitive dreams, and ESP [Abstract]. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1974* (pp. 47–48). Metuchen, NJ: Scarecrow Press.

*Haraldsson, E. (1980). Scoring in a precognition test as a function of the frequency of reading on psychical phenomena and belief in ESP. *Research Letter, 1,* 1–8.

Haraldsson, E. (1993). Are religiosity and belief in an afterlife better predictors of ESP performance than belief in psychic phenomena?. *Journal of Parapsychology*, *57*, 259–273.

*Haraldsson, E., Houtkooper, J. M., Schneider, R., & Bäckström, M. (2002). Perceptual defensiveness and ESP performance: Reconstructed DMT ratings and psychological correlates in the first German DMT-ESP experiment. *Journal of Parapsychology, 66,* 249–270.

*Hitchman, G. A., Pfeuffer, C. U., Roe, C. A., & Sherwood, S. J. (2016). The effects of experimenter-participant interaction qualities in a goal-oriented nonintentional precognition task. *Journal of Parapsychology*, *80*, 45–69.

*Hitchman, G. A., Roe, C. A., & Sherwood, S. J. (2012). A reexamination of nonintentional precognition with openness to experience, creativity, psi beliefs, and luck beliefs as predictors of success. *Journal of Parapsychology*, *76*, 109–145.

*Hitchman, G. A., Roe, C. A., & Sherwood, S. J. (2015). The relationship between lability and performance at inten-

tional and nonintentional versions of an implicit PMIR-type psi task. *Journal of Parapsychology, 79*, 65–86.
*Hitchman, G. A., Sherwood, S. J., & Roe, C. A. (2015). The relationship between latent inhibition and performance at a non-intentional precognition task. *Explore: The Journal of Science and Healing*, *11*, 118–126. doi:10.1016/j.explore.2014.12.004
Holt, N. J. (2002). *Creativity and boundary permeability*. Unpublished manuscript, University of Northampton, UK.
*Honorton, C. (1972). Reported frequency of dream recall and ESP. *Journal of the American Society for Psychical Research*, *66*, 369–374.
Honorton, C., & Ferrari, D. C. (1989). "Future telling": A meta-analysis of forced-choice precognition experiments, 1935-1987. *Journal of Parapsychology*, *53*, 281–308.
Honorton, C., Ferrari, D. C., & Bem, D. J. (1998). Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology*, *62*, 255–276.
*Humphrey, B. M. (1945). An exploratory correlation study of personality measures and ESP scores. *Journal of Parapsychology*, *9*, 116–123.
Hyman, R. (1977). The case against parapsychology. *The Humanist, 37*, 37–49.
Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49*, 3–49.
Hyman, R. (2010). Meta-analysis that conceals more than it reveals: Comment on Storm et al. (2010). *Psychological Bulletin, 136*, 486–490. doi:10.1037/a0019676
Kennedy, J. E. (2001). Why is psi so elusive? A review and proposed model. *Journal of Parapsychology, 65*, 219–246.
Kennedy, J. E. (2016). Is the methodological revolution in psychology over or just beginning? *Journal of Parapsychology*, 80, 156–168.
Knapp, R. H., & Garbutt, J. T. (1958). Time imagery and the achievement motive. *Journal of Personality*, *26*, 426–434. doi:10.1111/j.1467-6494.1958.tb01597
Kragh, U., & Smith, G. W. (Eds.) (1970). *Percept genetic analysis*. Lund, Sweden: Gleerup.
Lawrence, T. R. (1993). Gathering in the sheep and goats: A meta-analysis of forced-choice sheep-goat ESP studies, 1947-1993. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention,* 75–86.
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Götzsche, P. C., Ioannidis, J. P. A., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ, 339*, b2700. doi:10.1136/bmj.b2700.
Luke, D. P., Delanoy, D., & Sherwood, S. J. (2003). *Questionnaire of beliefs about luck*. Unpublished instrument, University of Northampton, UK.
*Luke, D. P., Delanoy, D., & Sherwood, S. J. (2008). Psi may look like luck: Perceived luckiness and beliefs about luck in relation to precognition. *Journal of the Society for Psychical Research, 72*, 193–207.
*Luke, D. P., & Morin, S. (2009). *Luck beliefs, PMIR, psi and the sheep-goat effect: A replication.* Abstracts of papers presented at the 33rd International Conference of the Society for Psychical Research, University of Nottingham (pp. 12–13).
*Luke, D. P., Roe, C. A., & Davison, J. (2008). Testing for forced-choice precognition using a hidden task: Two replications. *Journal of Parapsychology, 72*, 133–154.
Mangan, G. L. (1958). *A review of published research on the relationship of some personality variables to ESP scoring level*. New York, NY: Parapsychology Foundation.
McCarthy, D., & Schechter, E. I. (1986). Estimating effect size from critical ratios [Abstract]. In D. H. Weiner & D. I. Radin (Eds.), *Research in parapsychology 1985* (pp. 95–96). Metuchen, NJ: Scarecrow Press.
McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology, 52*, 81–90. doi:10.1037/0022-3514.52.1.81
Milton, J., & Wiseman, R. (1999). Does Psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, *125*, 387–391. doi:10.1037/0033-2909.125.4.387
Mousseau, M. C. (2003). Parapsychology: Science or pseudo-science. *Journal of Scientific Exploration*, *17*, 271–282.
Murray, A. L. (2011). The validity of the meta-analytic method in addressing the issue of psi replicability. *Journal of Parapsychology*, *75*, 261–277.
Nazimuddin, S. K. (2015). A study of individual differences in educational situations. *International Journal of Scientific Engineering and Research, 3*, 180–184.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349,* aaac4716-1–aaac4716-8. doi:10.1126/science.aac4716
Palmer, J. (1978). Extrasensory perception: Research findings. In S. Krippner (Ed,), *Advances in parapsychological research 2* (pp. 59–243). New York, NY: Plenum. doi:10.1007/978-1-4615-9092-7_3

Palmer, J. (2003). ESP in the ganzfeld: Analysis of a debate. *Journal of Consciousness Studies*, *10*, 51–68.

*Palmer, J. (2009). Decision augmentation in a computer guessing task. J*ournal of Parapsychology, 73*, 119–136.

Palmer, J. (2016). Hansel's ghost: Resurrection of the experimenter fraud hypothesis in parapsychology [Editorial]. *Journal of Parapsychology*, *80*, 5–16.

Persinger, M. A., & Makarec, K. (1987). Temporal lobe epileptic signs and correlative behaviors displayed by normal populations. *Journal of General Psychology, 114,* 179–195. doi:10.1080/00221309.1987.9711068

Rhine, J. B., Pratt, J. G., Stuart, C. E., Smith, B. M., & Greenwood, J. A. (1940). *Extra-sensory perception after sixty years: A critical appraisal of the research in extra-sensory perception*. New York, NY: Henry Holt. doi:10.1037/13598-000

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641. doi:10.1037/0033-2909.86.3.638

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, *4*, 61–81. doi:10.1007/s11292-007-9046-9

Schmeidler, G. R. (1943). Predicting good and bad scores in a clairvoyance experiment. A preliminary report. *Journal of the American Society for Psychical Research*, *37*, 103–110.

*Schmeidler, G. R. (1964a). An experiment on precognitive clairvoyance. Part I. The main results. *Journal of Parapsychology*, *28*, 1–14.

*Schmeidler, G. R. (1964b). An experiment on precognitive clairvoyance, Part III. Precognition scores related to the subject's way of viewing time. *Journal of Parapsychology, 28*, 94–101.

*Schmeidler, G. R. (1964c). An experiment on precognitive clairvoyance: Part IV. Precognition scores related to creativity. *Journal of Parapsychology, 28*, 102–108.

*Schmeidler, G. R. (1964d). An experiment on precognitive clairvoyance: Part V. Precognition scores related to feelings of success. *Journal of Parapsychology, 28*, 109–125.

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. doi:10.1348/000712604773952449

Smith, M. D. (1998). *Perception of one's own luck: The formation, maintenance and consequences of perceived luckiness* (Unpublished doctoral thesis). University of Hertfordshire, UK.

*Stanford, R. G. (1970). Extrasensory effects upon memory. *Journal of the American Society for Psychical Research*, *64*, 161–186.

Steinkamp, F. (1998). Psychology, psi, and the web: An exploratory study. *Journal of the American Society for Psychical Research, 92*, 256–278.

Steinkamp, F. (2005). Forced-choice ESP experiments: Their past and their future. In M. A. Thalbourne & L. Storm (Eds.), *Parapsychology in the 21st century: Essays on the future of psychical research* (pp. 124–163). Jefferson, 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*, 193–218.

*Storm, L. (2002). A parapsychological investigation of the I Ching: Seeking psi in an ancient Chinese system of divination. *Australian Journal of Parapsychology, 2*, 44–62.

Storm, L. (2006). Technical Paper No. 11: Meta-analysis in parapsychology: I. The ganzfeld domain. *Australian Journal of Parapsychology*, *6*, 35–53.

*Storm, L. (2008). Investigations of the I Ching: I. Relationships between psi and time perspective, paranormal belief and meaningfulness. *Australian Journal of Parapsychology*, *8*, 103–127.

*Storm, L., & Thalbourne, M. A. (1998-1999). The transliminal connection between personality and paranormal effects in an experiment with the I Ching. *European Journal of Parapsychology, 14*, 100–124.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2012). Meta-analysis of ESP studies, 1987-2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology*, *76*, 243–273.

Tart, C. T. (1983). Information acquisition rates in forced-choice ESP experiments: Precognition does not work as well as present-time ESP. *Journal of the American Society for Psychical Research, 77*, 293–310.

*Thalbourne, M. (1996). An attempt to predict precognition scores using transliminality-relevant variables. *Journal of the Society of Psychical Research, 62*, 129–140.

Thalbourne, M. A. (1998). Transliminality: Further correlates and a short measure. *Journal of the American Society for Psychical Research, 92*, 402–419.

Thalbourne, M. A., & Delin, P. S. (1993). A new instrument for measuring the sheep-goat variable: Its psychometric properties and factor structure. *Journal of the Society for Psychical Research, 59*, 172–186.

Thalbourne, M. A., & Delin, P. S (1994) A common thread underlying belief in the paranormal, mystical experience

and psychopathology. *Journal of Parapsychology, 58*, 3–38.
*Thalbourne, M. A., & Storm, L. (2014). A further study of psychopraxia using the "I Ching." *Australian Journal of Parapsychology*, *14*, 115–142.
Torrance, E. P. (1963). *Preliminary Manual for Personal-Social Motivation Inventory*. Mineapolis, MN: University of Minnesota.
Utts, J. (1991). Replication and meta-analysis in parapsychology. *Statistical Science, 6*, 363–378.
Utts, J. (2015). What constitutes replication in parapsychology?. In E. C. May, & S. B. Marwaha (Eds.), *Extrasensory perception: Support, skepticism, and science* (Vol. 1). Santa Barbara, CA: ABC-CLIO.
Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics, 35*, 215–247. doi:10.3102/1076998609346961
*Vaughn, A., & Houck, J. (1993). A "success" test of precognition and attitude toward the future. *Journal of the American Society for Psychical Research*, *59*, 259–268.
Watt, C. (2005). Parapsychology's contribution to psychology: A view from the front line. *Journal of Parapsychology*, *69*, 215–231.
Watt, C. (2016). A prospective meta-analysis of pre-registered ganzfeld ESP studies. KPU Study Registry. Retrieved from http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_Registry_1024.pdf
Watt, C., & Kennedy, J. E. (2015). Lessons from the first two years of operating a study registry. *Frontiers in psychology*, *6*, 173. doi:10.3389/fpsyg.2015.00173
Watt, C., & Kennedy, J. E. (2016). Options for prospective meta-analysis and introduction of registration-based prospective meta-analysis. *Frontiers in Psychology, 7*, 2030. doi:10.3389/fpsyg.2016.02030
*Winkelman, M. (1981). The effect of formal education on extrasensory abilities: The Ozolco study. *Journal of Parapsychology*, *45*, 321–336.
Wiseman, R., & Greening, E. (2002). The Mind Machine: A mass participation experiment into the possible existence of extrasensory perception. *British Journal of Psychology*, *93*, 487–499. doi:10.1348/000712602761381367

Meta-Analysis Papers Not Cited

Haraldsson, E., & Johnson, M. (1979). ESP and the Defense Mechanism Test (DMT) Icelandic study No. III: A case of the experimenter effect? *European Journal of Parapsychology, 3*, 11–20.
Haraldsson, E. (1978). ESP and the defense mechanism test (DMT): A further validation. *European Journal of Parapsychology*, *2*, 104–114.
Haraldsson, E., & Johnson, M. (1986). The Defense Mechanism Test (DMT) as a predictor of ESP performance: Icelandic studies VI and VII [Abstract]. In D. H. Weiner & D. I. Radin (Eds.), *Research in parapsychology 1985* (pp. 43–44). Metuchen, NJ: Scarecrow Press.
Honorton, C. (1967). Creativity and precognition scoring level. *Journal of Parapsychology*, *31*, 29–42.
Johnson, M. (1969). Attitude and target differences in a group precognition test. *Journal of Parapsychology*, 33, 324–325.
Johnson, M., & Haraldsson, E. (1984). The Defense Mechanism Test as a predictor of ESP scores: Icelandic studies IV and V. *Journal of Parapsychology*, *48*, 185–200.
Nielsen, W. (1970). Relationships between precognition scoring level and mood. *Journal of Parapsychology*, *34*, 93–116.
Ryzl, M. (1968a). Precognition scoring and attitude toward ESP. *Journal of Parapsychology*, *32*, 1–8.
Ryzl, M. (1968b). Precognition scoring and attitude. *Journal of Parapsychology*, *32*, 183–189.
Schwartz, S. A., & De Mattei, R. J. (1983). The Mobius Psi-Q Test: Preliminary findings [Abstract]. In R. A. White & J. Solfvin (Eds.), *Research in parapsychology 1984* (pp. 103–105). Metuchen, NJ: Scarecrow Press.
Smith, M. D., Wiseman, R., Machin, D., Harris, P., & Joiner, R. (1997). Luckiness, competition and performance on a psi task. *Journal of Parapsychology, 61*, 33–43.
Storm, L. (2006). A parapsychological investigation of the I Ching: The relationship between psi, intuition, and time perspective. *Journal of Parapsychology*, *70*, 121–141.
Storm, L., & Thalbourne, M. A. (2001). Studies of the I Ching: I. A replication. *Journal of Parapsychology, 65*, 105–124.
Thalbourne, M., Beloff, I., & Delanoy, D. (1982). A test for the "extraverted sheep versus introverted goats" hypothesis [Abstract]. In W. G. Roll, R. L. Morris, & R. A. White (Eds.), *Research in parapsychology 1981* (pp. 155–156). Metuchen, NJ: Scarecrow Press.
Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely

confirmatory research. *Perspectives on Psychological Science*, *7*, 632–638. doi:10.1177/1745691612463078
Wilson, S., & Hamlin, I. (2007). Implicit learning in a card prediction task. *European Journal of Parapsychology, 22*, 3–29.

Acknowledgements

The authors would like to thank Caroline Watt, Gergö Hadlaczky, and Jan Dalkvist for their comments on an earlier draft of this paper. We would also like to thank John Palmer and the two anonymous referees for their constructive feedback.

Victoria University of Wellington
Kelburn Parade
PO Box 600
Wellington 6012, New Zealand
Marco.Zdrenka@vuw.ac.nz

Appendix

Table A1
Number of Studies for Each Individual Difference Measure That Reported Nonsignificance Without Providing Any Other Statistical Information

Individual Difference Measure	Number of non-significant studies estimated as $r = .00$
Belief in Psi	5
Extraversion	3
Neuroticism	2
Agreeableness	2
Spontaneous Experiences	2
Openness to Experience	1
Dream Recall	1
Conscientiousness	1
Temporal Lobe Dysfunction	1
Cerebral Lateralization	1

Table A2
68% and 80% Confidence Intervals for All Individual Difference Measures Meta-Analysed

Individual Difference Measure	68% Confidence Intervals	80% Confidence Intervals
Chance Belief	.03, .25	-.01, .28
Luck Belief	.07, .20	.05, .21
Belief in Psi	.09, .16	.07, .18
Perceptual Defensiveness	.06, .19	.04, .20
Providence Belief	.05, .20	.02, .22
Openness to Experience	.07, .17	.06, .19
Extraversion	.05, .11	.04, .12
Perceived Luckiness	-.03, .18	-.06, .21
Dream Recall	.01, .13	-.004, .14
Intelligence	-.19, .07	-.23, .10
Emotional Reactivity	-.11, .23	-.16, .28
Conscientiousness	-.02, .13	-.04, .15
Neuroticism	-.01, .12	-.03, .14
Religiosity	-.15, .05	-.18, .07
Future-Oriented	-.04, .15	-.07, .17
Creativity	-.02, .11	-.04, .13
Fortune Belief	-.03, .10	-.05, .12
Agreeableness	-.04, .08	-.05, .10
Pro Attitude	-.07, .03	-.08, .05
Time Belief as Dynamic	-.03, -.01	-.03, -.01
Transliminality	-.06, .08	-.08, .10
Spontaneous Experiences	-.04, .03	-.05, .04
Temporal Lobe Dysfunction	-.10, .09	-.13, .12

Abstracts in Other Languages

French

LES CORRELATS DES DIFFERENCES INTERINDIVIDUELLES DES PERFORMANCES PSI DANS LES EXPERIMENTATIONS DE PRECOGNITION A CHOIX FORCE : UNE META-ANALYSE (1945-2016)

RESUME : Les précédentes recherches en parapsychologie ne se sont pas révélées particulièrement convaincantes, en grande partie du fait de manque de réplicabilité des résultats significatifs. Pour traiter ce problème et mieux comprendre les facteurs qui pourraient être associés avec des tailles d'effet plus fortes et consistantes, toutes les expérimentations de précognition à choix forcé analysant des différences interindividuelles (par exemple, des traits de personnalité) ont été rassemblées pour déterminer quels facteurs peuvent prédire de façon fiable les performances psi. Globalement, un ensemble de 55 études publiées entre 1945 et 2016, comprenant 35 mesures des différences interindividuelles, fut soumis à une méta-analyse. Six mesures de différence interindividuelle se sont révélées être corrélées significativement avec les performances psi, à savoir : la croyance à la chance (croyance que la chance peut être directement contrôlée), la défense perceptive, l'ouverture à l'expérience, la croyance au psi, l'extraversion et la croyance à un temps dynamique. Etant donnée la nature particulièrement directe des expérimentations de précognition à choix forcé, il semble prometteur de continuer à explorer ces facteurs dans des études confirmatoires. Nous espérons que des chercheurs pourront modeler leurs futures expérimentations en conjonction avec des techniques de pré-enregistrement, afin de créer ultimement une base de données plus systématique et robuste.

German

KORRELATE INTERINDIVIDUELLER UNTERSCHIEDE VON PSI-LEISTUNGEN BEI PRÄKOGNITIONSEXPERIMENTEN MIT BEGRENZTER WAHL: EINE META-ANALYSE (1945-2016)

ZUSAMMENFASSUNG: Die bisherige Forschung in der Parapsychologie war nicht besonders überzeugend, was größtenteils mit der mangelnden Replizierbarkeit signifikanter Ergebnisse zusammenhängt. Um diesem Mangel abzuhelfen und um besser zu verstehen, welche Faktoren zu stärkeren und konsistenteren Effektstärken beitragen könnten, wurden alle Präkognitionsexperimente mit begrenzter Wahl, die interindividuelle Unterschiede untersuchten, z. B. Persönlichkeitsmerkmale, zusammengefasst, um zu bestimmen, welche Faktoren zuverlässig Psi-Leistungen vorhersagen könnten. Insgesamt wurden 55 Studien, die zwischen 1945 und 2016 veröffentlicht worden waren und 35 Messgrößen für interindividuelle Unterschiede verwendeten, für eine Meta-Analyse zusammengefasst. Bei sechs Messgrößen für interindividuelle Unterschiede, nämlich Glaube an Glück (der Glaube, dass Glück primär kontrollierbar sei), Wahrnehmungsabwehr, Offenheit für Erfahrungen, der Glaube an Psi, Extraversion und der Glaube, dass Zeit dynamisch sei, zeigte sich eine signifikante Korrelation mit der Psi-Leistung. Angesichts der besonders einfachen Natur dieser Art von Experimenten könnte eine erfolgversprechende Forschungsrichtung darin bestehen, diese Faktoren in Bestätigungsexperimenten zu untersuchen. Es wäre zu hoffen, dass die Forscher bei ihren zukünftigen Experimenten diese Ergebnisse berücksichtigen, um – zusammen mit den Techniken der vorherigen Registrierung – zur Schaffung einer letztlich systematischeren und robusteren Datenbank beizutragen.

Spanish

CORRELATOS DE DIFERENCIAS INDIVIDUALES EN EXPERIMENTOS DE PRECOGNICIÓN DE ELECCIÓN FORZADA: UN META-ANÁLISIS (1945-2016)

RESUMEN: Las investigaciones previas en parapsicología no han sido particularmente convincentes, en gran parte debido a la falta de replicabilidad de los hallazgos significativos. Para resolver este problema y comprender mejor qué factores pueden estar asociados con tamaños de efecto más fuertes y consistentes, se agregaron todos los experimentos de precognición de elección forzada con análisis de diferencias individuales (p. ej., rasgos de personalidad) para determinar qué factores podrían predecir fiablemente el rendimiento psi. En total, 55 estudios publicados entre 1945 y 2016, incluyendo 35 medidas de diferencia individuales, fueron objeto de meta-análisis. Seis medidas de diferencias individuales correlacionaron significamente con la tarea psi: la creencia en la suerte (la creencia de que la suerte es básicamente controlable), defensividad perceptual, apertura a la experiencia, creencia en psi, extraversión, y creencia en un tiempo dinámico. Dada la naturaleza clara de los experimentos de precognición de elección forzada, una vía futura prometedora sería explorar estos factores en estudios confirmatorios. Se espera que los investigadores puedan modelar sus futuros experimentos en estos hallazgos en conjunción con las técnicas de prerregistro, para crear una base de datos más sistemática y robusta.

Journal of Parapsychology, 81(1), 33-45

BEM'S "FEELING THE FUTURE" (2011) FIVE YEARS LATER: ITS IMPACT ON SCIENTIFIC LITERATURE

By Bruno A. Silva and Gabrielle Poeschl

ABSTRACT: The study analyses the impact on scientific literature of the controversial 2011 article by Bem, "Feeling the Future." Texts that cite Bem´s article (*N* = 162) were identified from the Elsevier Scopus database for the years 2011 to 2015. Their summaries were analyzed using the Iramuteq program for textual data. The analysis suggested that the impact can be grouped into four classes: (a) The Replication class is characterized by a vocabulary addressing the role of replication in psychology research; (b) the Bayesian class reflects the perceived merits of this approach when compared with more traditional inferential statistics, namely statistics relying on *p* values; (c) the Experimental Studies of Anomalous Experiences (AE) class includes terms related to classical empirical research, applied in AE studies, and concepts, methodologies and theories specific to parapsychological research; (d) the Quantum Phenomena and Theories class suggests that quantum theories of brain/consciousness function may leave the door open to the possibility of the existence of psi phenomena. The Replication and Bayesian classes are represented more in psychology literature, confirming our hypothesis that psychology sources have a more critical position. Data moreover suggest that replication is currently the most referred path in the attempt to reach closure on this controversy.

Keywords: scientific controversy, precognition, retrocausation, Daryl Bem, parapsychology

Even before its publication, in March 2011, in the highly influential *Journal of Personality and Social Psychology (JPSP)*, Daryl Bem´s (2011) article "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect" had already initiated an intense debate in the (social) media, establishing a remarkable controversy (e.g., Gad-el-Hak, 2011; MSNBC, 2011). The main reason for this was the message conveyed by its findings: "Precognition is scientifically supported" (Duggan, 2015). Although there is great public interest in parapsychological (psi) research, controversy has always been an integral part of this field of study, such that Richard Broughton (1991) titled his book *Parapsychology: The Controversial Science*. Evidence for psi was reported in the extensive parapsychological literature of recent and past years. What, then, made the article by Bem so unique and interesting? According to Palmer (2015), its particular importance lies in the scientific status of Bem as a social psychologist at the prestigious Cornell University, where he works, and in the fact that in this article he adapted experiments from methodological paradigms that are common in psychology, although reversing the order in which the task stimuli were presented.

Although controversies in science are common and scientists often argue among themselves, this does not mean that scientists must argue or that controversies are essential to the development of science. The history of science shows that some important theoretical changes have been controversial (e.g., Darwinism), whereas other achievements (e.g., electricity-magnetism unification) have not (Machamer, Pera, & Baltas, 2000). There are also controversies that are explicitly avoided. By publishing in a highly regarded journal, and using well established methodological paradigms of mainstream psychology that were reversed, Bem also gave rise to controversy within the scientific community.

In parapsychology, some controversies have been documented and studied. Zingrone (2006) presents an impressive amount of data and an in-depth analysis of controversies throughout its history. Palmer (2015) analyzed Bem´s (2011) controversy in some depth, namely the specific criticisms levelled at his studies and the meta-analysis conducted by Tressoldi, Rabeyron, Duggan, and Bem (2016; see also Bem, Tressoldi, Rabeyron, and Duggan, 2015), which supports the early findings of Bem.

The present study aims to complement this more in-depth approach to the controversies in parapsychology through an analysis of texts indexed in a scientific database that cited Bem´s (2011)

article, using the Iramuteq program for textual data analysis (Ratinaud, 2009). Using exploratory means, Iramuteq employs a quantitative approach that we used in order to understand the impact of this article on the scientific community.

We analyzed three supplementary variables representing the year of publication and subject area. The choice of the supplementary variables resulted from our expectations. Indeed, we did not have any particular predictions regarding the year of publication, but we had some hypotheses with respect to the subject areas in which the sources were indexed: We expected that a more critical point of view on Bem's article would be demonstrated by papers published in sources indexed in the subject area Psychology, and a more positive point of view would be found in texts published in sources indexed in other subject areas (see, e.g., McClenon, 1982; McClenon, Roig, Smith & Ferrier, 2003; McConnell & Clark, 1991; Roe, 2016).

Method

Data Collection

The texts that cite Bem´s (2011) "Feeling the Future" were identified from the Elsevier Scopus database for the years 2011 to 2015. The export procedure was performed on August 24, 2015, resulting in the collection of 163 texts, but one that was duplicated was excluded from further analysis. The following information was registered for the 162 remaining texts: (a) title, (b) abstract (when available), (c) authors' keywords (when available), (d) year of publication, (e) authors' names, (f) source, and (g) subject area. A source could be indexed in more than one subject area. Very few errors were detected in the downloaded texts (eight words misspelled, two authors' keywords missing and one incomplete). These errors were manually corrected after consulting the original texts, and one title was translated from Italian to English in order to run the automatic data analysis on a reliable document. Titles, abstracts, authors' keywords, years of publication, and authors' names were then compiled into a single textual *corpus,* which constituted the material to be analyzed.

Summaries were separated by the supplementary variables necessary for the association of the lexical productions with the years and subject areas of publication. Three supplementary variables were added: first, the year of publication [variable with five modalities (or categories): from 2011 to 2015], and then two different (but complementary) variables for the definition of the subject areas (e.g., Psychology, Medicine) in which the sources (academic journals, books, conference proceedings) were indexed: (a) a variable with two modalities: in the subject area Psychology, indexation (i) or not (ii), and (b) a variable with three modalities: indexation only in the subject area Psychology (i), or indexation in the subject area Psychology as well as in other subject areas, namely, Arts and Humanities, Medicine, Social Sciences, Mathematics, Neuroscience, Business, Management and Accounting, Decision Sciences, Economics, Econometrics and Finance (ii), or indexation *not* in subject area Psychology but in other subject areas, which were, in addition to the above-mentioned ones, Physics and Astronomy, Multidisciplinary, Nursing, Agricultural and Biological Sciences, Computer Science, Engineering, Materials Science, and Health Professions (iii). The introduction of this second variable for the subject areas in which the sources were indexed was meant to assess whether the lexical production of the texts included in sources indexed in subject area Psychology and also in other subject areas would modulate the broad difference expected between the texts indexed or not indexed in subject area Psychology.

Data Analysis

To analyze the very large corpus extracted from the database, we used the Iramuteq program (Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires). This program, anchored in R software, enables several different statistical analyses and processing of text (Ratinaud, 2009). In particular, Iramuteq enables the analysis of a corpus using a method developed by Reinert (e.g., 1993), named Alceste (Analyse des Lexèmes Cooccurrents dans les Énoncés Simples d'un Texte), which

has been available for a long time in the software ALCESTE®. The method enables the study of the formal structure of the co-occurrence of the words ("*the lexical worlds"*) in a given text or set of texts: Classes of words are created and associated with supplementary variables likely to explain the conditions of their production. More precisely, the program performs a downward hierarchical classification, extracting (creating) classes of words that co-occur and that are most dissimilar from other classes of co-occurring words; it presents the resulting clustering tree (*dendrogram*) showing the proximity relationships between the classes; it also provides a list of the most representative words in each class, a list of text segments in which the representative words are included, and the modalities (i.e., the categories) of the supplementary variables associated with the classes. Iramuteq also performs an analysis of correspondence (Benzécri, 1973), which provides a schematic spatial representation of the relationships between the extracted classes, represented by their most representative words. Data analysis was performed using Iramuteq version 07.2, released on December 23, 2014.

Concretely, three steps are taken to perform the analyses. In Step 1 the different parts of a text or, in the case of a questionnaire, respondents' answers to an open-ended question (see, e.g., Poeschl, Valentim, & Silva, 2015), are divided into text segments. The text segments correspond, usually, to a sentence and they are the basis of all computed statistics. At that time, all words are registered to create a dictionary, as well as a dictionary of their lemmatized *forms* (words reduced to their roots), intended to avoid irrelevant differences (words written in singular or plural, verbs conjugated in different tenses, etc.; Chabchoub, 2008). The text segments are congregated in larger groups of text segments which form the rows of a table. The lemmatized forms form the columns of the table. The intersection of the rows and the columns registers either 1 if the form is present in the text segments or 0 if it is not.

In Step 2, the table is submitted to a downward hierarchical classification, which iteratively separates the text segments into classes in which specific lemmatized forms co-occur and classes in which they do not (Reinert, 1986). This procedure was developed to analyze large logical tables (codified 0 or 1) of scarce frequencies. The first iteration decomposes the original table, made up of all the text segments, into two subtables, one including text segments with mostly 0 in the cells in the original table, and the other including text segments with mostly 1 in the cells. Each of the subtables thus constitutes a class made up of a distinct vocabulary with the overlap of classes as small as possible. The following iterations further decompose the subclasses, increasing the intraclass homogeneity and the interclass differences until the iterations fails to improve the sorting (Reinert, 1986; for more details about the procedure, see also Reinert, 1983, 1985). The procedure results in a tree diagram.

In Step 3, the results that can be interpreted are displayed. The classes presented in the clustering tree are described by the list of their associated forms, which constitutes the classes' *profiles*. The profile of a class is the list of the words associated with the class. The degree of association of a form with a class is assessed by applying the chi-square formula on the observed data. An illustration of the relation of Form 1 (Replication) with Class 1 (The Role of Replication in Psychology Research) is presented in Table 1 (Reinert, 1993; see Pélissier, 2016, for further examples). In this example, the chi square of association is significant: $\chi^2(1, N = 622) = 105.27, p < .0001$.

Table 1
Presence Versus Absence of Form 1 in the Text Segments of Class 1 Versus Other Classes in the Observed Data

	Number of text segments with Form 1 present	Number of text segments with Form 1 absent	Total
Class 1	55	140	195
Other classes	7	420	427
Total	62	560	622

To indicate the direction of the relation between a form and a class (in Table 1, e.g, whether there were, or not, more Form 1 present in Class 1 than in the other classes, or vice versa), the sign of the difference between the observed and the expected frequencies is added to the chi square value, which is then denominated *chi square of association*. The significance of the chi square value, with one degree of freedom, is assessed and the classes are described by their most representative forms. The larger the chi square value, the stronger the association of the form with the class, and each class is thus characterized by a set of forms (words) more representative of that class than the others (Viaud, Uribe, & Acosta, 2007).

A chi-square of association is also computed to assess the degree of association of the supplementary variables introduced in the corpus with the classes extracted by the program. An illustration of the relation of Modality 2 (year 2012) of Supplementary Variable 1 (Year of Publication) with Class 2 (Experimental Studies of Anomalous Experiences) is presented in Table 2. The chi square of association is significant: $\chi 2(1, N = 622) = 22.34, p < .0001$.

Table 2
Presence Versus Absence of Modality 2 of Supplementary Variable 1 in the Text Segments of Class 2 Versus Other Classes in the Observed Data

	Number of text segments with Modality 2 present	Number of text segments with Modality 2 absent	Total
Class 2	63	90	153
Other classes	102	367	469
Total	165	457	622

We present hereafter the results obtained from the procedure. More specifically, we describe the profiles of the classes extracted by the downward hierarchical classification performed on the corpus, and their association with the modalities of the supplementary variables. Then we present the results of the correspondence analysis.

Results

The analysis of the corpus with the Iramuteq program was performed on the 162 texts that cited Bem´s (2011) "Feeling the Future." In step 1, the corpus was subdivided by the program into 721 text segments containing, altogether, 4438 lemmatized forms. In step 2, the downward hierarchical classification sorted 622 text segments into four classes, classifying 86.27% of the 721 total text segments. The excluded segments contained forms (i.e., words) that were too rare to be taken into consideration (i.e., total frequency less than 3). The analyzed text segments were first divided into two main *branches* of the clustering tree (or tree diagram) on the basis of the dissimilarity of their vocabulary. Each one was subsequently decomposed further into two *classes* of words that did not require any further division (see Figure 1).

First Branch of the Dendrogram

We begin by describing the two classes of the first branch of the dendrogram presented in Figure 1, which includes 58.3% of the total text segments: the large Class 1 and the smaller Class 4. Table 3 presents a summary of the forms included in these two classes, that is, their 35 most representative forms, showing the frequency of each form in the class, its total frequency in the analyzed segments, the percentage of the form in the class, and the value of its chi square of association, calculated from a table such as Table 1 (for each value, $df = 1$; $N = 622$, $p < .001$). To make the reading easier, we substituted for each lemmatized form the word most frequently used in the corpus.

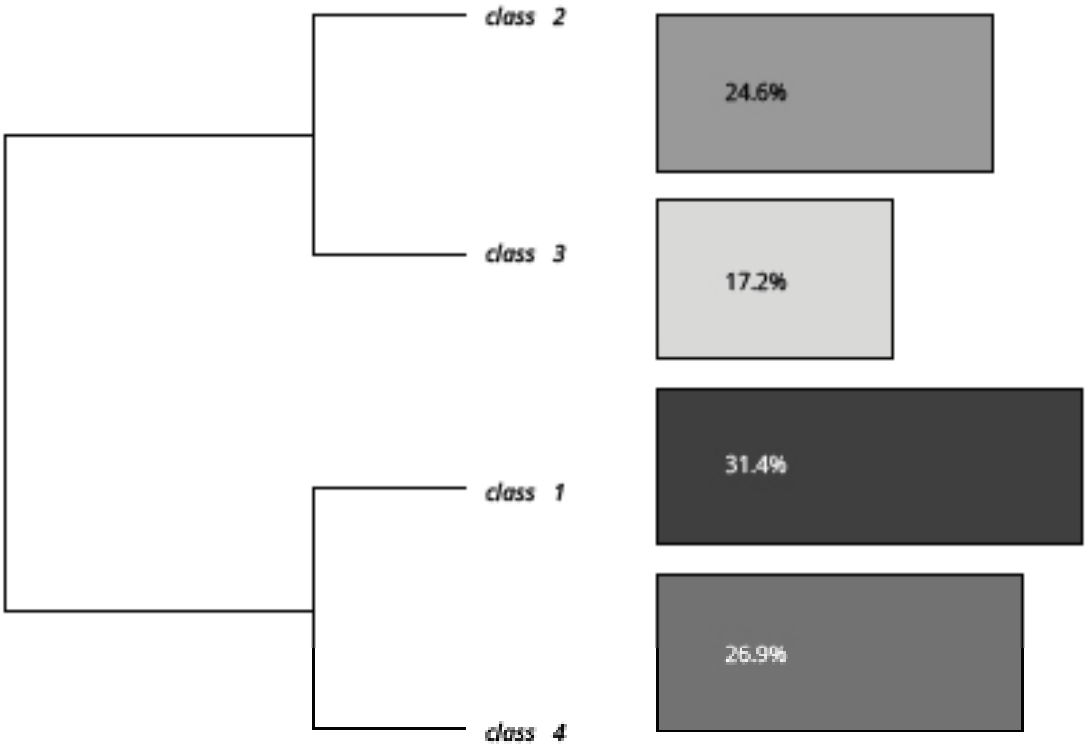


Figure 1. Dendrogram and relative weights of the classes based on the percentage of text segments.

Class 1 - The Role of Replication in Psychology Research. Class 1, one of the two largest classes, is composed of 195 text segments, that is, 31.4% of the corpus, organized around "replication": $\chi^2 = 105.27$. As may be seen in Table 3, it also includes significant chi squares for the words "psychology" ($\chi^2 = 88.38$),"science" ($\chi^2 = 76.15$), "research" ($\chi^2 = 72.64$), "article" ($\chi^2 = 37.84$), "publication bias" ($\chi^2 = 31.99$), "practice" ($\chi^2 = 31.76$), "psychological" ($\chi^2 = 27.72$), "important" ($\chi^2 = 27.44$), "method" ($\chi^2 = 26.28$), "empirical" ($\chi^2 = 25.96$), "education" ($\chi^2 = 25.18$), "field" ($\chi^2 = 22.72$), "publication" ($\chi^2 = 20.68$), "methodological" ($\chi^2 = 19.67$), "recommendation" ($\chi^2 = 18.46$), "improve" ($\chi^2 = 16.90$), "credibility" ($\chi^2 = 15.50$), "validity" ($\chi^2 = 13.27$), "crisis" ($\chi^2 = 13.25$), and "error" ($\chi^2 = 11.87$). This class is characterized by a vocabulary that addresses the role of replication in psychology research. The lack of replication of some studies, along with questionable research practices, may have contributed to a crisis of credibility regarding psychological science. Publication bias is a major concern, and the perceived crisis is also viewed as an opportunity to improve research practices through implementation of recommendations that can increase the credibility of scientific evidence in psychology journals. Issues of validity of empirical findings and errors also characterize the vocabulary of this class.

This first class groups vocabulary that was referred to mostly in texts whose sources are indexed in subject area Psychology ($f = 147$, $\chi^2 = 36.96$). Significant chi squares were found for only subject area Psychology ($f = 68$, $\chi^2 = 13.44$); Psychology and Other Subject Areas ($f = 79$, $\chi^2 = 9.1$); and texts published in 2015 ($f = 20$, $\chi^2 = 11.46$). The other chi squares are nonsignificant. With regard to the subject areas, we note that, although globally the vocabulary stems from texts whose sources are indexed in subject area Psychology, it is more specifically characteristic of the texts whose sources are indexed only in this area than of the texts whose sources are indexed in the modality Psychology and also other subject areas.

Table 3
Class profiles: 35 Most Representative Words Associated With Class 1 and Class

Class 1					Class 4				
Form	*N* in class	Total *N*	% in class	χ^2	Form	*N* in class	Total *N*	% in class	χ^2
Replication	55	62	88.71	105.27	Bayes	31	31	100.00	88.89
Psychology	62	81	76.54	88.38	Factor	35	42	83.33	73.17
Science	47	57	82.46	76.15	Bayesian	32	39	82.05	64.56
Research	61	86	70.93	72.64	Test	41	61	67.21	56.11
Scientific	31	40	77.50	42.30	Null	25	29	86.21	54.57
Article	30	40	75.00	37.84	Bayes factor	18	19	94.74	45.99
Publication bias	16	17	94.12	31.99	Statistical	27	36	75.00	45.11
Practice	31	45	68.89	31.76	Hypothesis	35	54	64.81	43.40
Psychological	24	33	72.73	27.72	Datum	42	71	59.15	42.59
Important	14	15	93.33	27.44	*p* values	16	17	94.12	40.27
Method	29	44	65.91	26.28	Inference	14	17	82.35	27.41
Empirical	21	28	75.00	25.96	Default	9	9	100.00	24.88
Education	13	14	92.86	25.18	Set	17	25	68.00	22.46
Field	16	20	80.00	22.72	Analyse	13	17	76.47	21.91
Publication	11	12	91.67	20.68	Provide	14	19	73.68	21.89
Methodological	12	14	85.71	19.67	Significance	10	12	83.33	19.88
Recommendation	10	11	90.91	18.46	Calorie	7	7	100.00	19.29
Core	8	8	100.00	17.75	Frequentist	7	7	100.00	19.29
Fact	8	8	100.00	17.75	Food	7	7	100.00	19.29
Improve	12	15	80.00	16.90	Analyze	8	9	88.89	17.90
Movement	9	10	90.00	16.24	Effect	31	63	49.21	17.84
Credibility	7	7	100.00	15.50	Obtain	9	11	81.82	17.23
Issue	12	16	75.00	14.54	Evidence	34	72	47.22	17.21
Replicability	11	14	78.57	14.84	Assessment	11	15	73.33	16.91
Journal	14	20	70.00	14.34	Effect size	11	15	73.33	16.91
Developmental	6	6	100.00	13.27	Probability	11	15	73.33	16.91
Guide	6	6	100.00	13.27	Stop	6	6	100.00	16.51
Validity	6	6	100.00	13.27	Yield	6	6	100.00	16.51
Crisis	9	11	81.82	13.25	Estimate	8	10	80.00	14.62
Investigation	9	11	81.82	13.25	Favor	8	10	80.00	14.62
Argue	16	25	64.00	12.90	Confirmatory	5	5	100.00	13.73
Error	7	8	87.50	11.87	Exploratory	5	5	100.00	13.73
Similar	7	8	87.50	11.87	One-sided	5	5	100.00	13.73
Finding	17	28	60.71	11.75	Small	5	5	100.00	13.73
Experimental	26	49	53.06	11.65	Value	12	19	63.13	13.16

Note. *N* = number. Critical values of χ^2 with 1 df: 19.51 for $p < .00001$: 15.14 for $p < .0001$; 10.83 for $p < .001$.

Class 4—Bayesian Statistical Inference. Close to Class 1 in terms of vocabulary, Class 4, the second largest class, includes 167 text segments which correspond to 26.9% of the total segments. This class is structured around "Bayes" (χ^2 = 88.89). As shown in Table 3, it also comprises significant chi squares for the words "factor" (χ^2 = 73.17), "Bayesian" (χ^2 = 64.56), "test" (χ^2 = 56.11), "null" (χ^2 = 54.57), "Bayes factor" (χ^2 = 45.99), "statistical" (χ^2 = 45.11), "hypothesis" (χ^2 = 43.40), "data" (χ^2 = 42.59), "*p* values" (χ^2 = 40.27), "inference" (χ^2 = 27.41), "default" (χ^2 = 24.88), "set" (χ^2 = 22.46), "analyse" (χ^2 = 21.91), "significance" (χ^2 = 19.88), "frequentist" (χ^2 = 19.29), "assessment" (χ^2 = 16.91), "effect size" (χ^2 = 16.91), "probability" (χ^2 = 16.91), "exploratory" (χ^2 = 13.73), and "one-sided" (χ^2 = 13.73). The vocabulary of this class points mainly to concepts related with the Bayesian approach. The segments of text mostly reflect the perceived merits of this approach when compared with the more traditional inferential statistics, namely those relying on *p* values. The words "Bayes" and "Bayesian" are used to express an inferential statistical method used to test hypotheses (null and alternative) in which the notion of probability is different from the frequentist approach. On the other hand, the specific term "Bayes factor" is a value that results from one Bayesian approach to assessing null values, and it is used to infer which model (null or alternative) is more accurate given the data. With regard to effect size, it relates to both approaches.

Class 4 groups vocabulary that is referred to more in texts whose sources are indexed in subject area Psychology (f = 121, χ^2 = 20.74); Psychology and Other Subject Areas (f = 78, χ^2 = 22.16); and Texts Published in 2011 (f = 51, χ^2 = 14.39).

Second Branch of the Dendrogram

The second branch of the dendrogram presented in Figure 1 includes 41.8% of the text segments, also divided into two classes of unequal weight, Class 2 and the small Class 3. Table 4 presents a summary of the forms included in these two classes, that is, their 35 most representative forms, indicating for each form its frequency in the class, its total frequency in the analyzed text segments, its percentage in the class, and the value of its chi square of association (df = 1, N = 622, p < .001). We also substituted for each lemmatized form the word most frequently used in the corpus, to make the reading easier.

Class 2—Experimental Studies of Anomalous Experience. Class 2 includes 153 text segments, that is, 24.6% of the segments. It is organized around "participant" and "task" ($\chi2$ = 43.10). It significantly comprises the words "experience" ($\chi2$ = 34.67),"target" ($\chi2$ = 31.15), "response" ($\chi2$ = 29.14), "predict" ($\chi2$ = 28.44), "learn" ($\chi2$ = 27.99), "anomalous" ($\chi2$ = 26.55), "nonintentional" ($\chi2$ = 24.84), "spontaneous" ($\chi2$ = 24.84), "stimulus" ($\chi2$ = 22.40), "performance" ($\chi2$ = 21.70), "precognition" ($\chi2$ = 20.43), "subject" ($\chi2$ = 19.68), "emotional" ($\chi2$ = 18.57), "prestimulus" ($\chi2$ = 18.57), "mind" ($\chi2$ = 18.19), "control" ($\chi2$ = 17.69), "randomly" ($\chi2$ = 17.29), "go/nogo" ($\chi2$ = 15.45), "PMIR" ($\chi2$ = 15.45), and "remote viewing" ($\chi2$ = 15.45). This class includes terms from classical empirical research, applied in anomalous experiences studies, and also concepts, methodologies, and theories specific to parapsychological research (e.g., non-intentional precognition task, prestimulus, go/nogo task, remote viewing, and psi-mediated instrumental response—PMIR).

Class 2 groups vocabulary that was referred to mostly in texts whose sources are not indexed in subject area Psychology (f = 88, $\chi2$= 18.87); and in Texts Published in 2012 (f = 63, $\chi2$ = 22.34).

Class 3—Quantum Phenomena and Theories. Class 3, with 107 text segments which correspond to 17.20% of the total analyzed segments, is the smallest of the extracted classes. It is significantly associated with "quantum" ($\chi2$ = 143.46), "brain" ($\chi2$ = 73.98), "consciousness" ($\chi2$ = 57.21), "Orch Or" ([Penrose–Hameroff theory of] orchestrated objective reduction; $\chi2$ = 39.01), "entanglement" ($\chi2$ = 37.81), "moment" ($\chi2$ = 37.81), "system" ($\chi2$ = 34.52), "retrocausality" ($\chi2$ = 34.08), "space time" ($\chi2$ = 34.08), "microtubule" ($\chi2$ = 29.16), "nonlocality" ($\chi2$ = 29.16), "superposition" ($\chi2$ = 29.16), "reality" ($\chi2$ = 28.12), "mechanic" ($\chi2$ = 24.26), "paa" (predictive anticipatory activity; $\chi2$ = 24.26), "causality" ($\chi2$ = 23.53), "principle" ($\chi2$ = 23.33), "anticipatory" ($\chi2$ = 19.38), "EKT" ([analysis of] Echeverria, Klinkhammer, and Thorne; $\chi2$ = 19.38), "paradox" ($\chi2$ = 19.38), "backward" ($\chi2$ = 18.60), and "bilk" ($\chi2$ = 18.60).

This last class grouped vocabulary that was mainly mentioned in texts whose sources are not indexed in subject area Psychology (f = 82, $\chi2$ = 61.84); and in Texts Published in 2011 (f = 34; $\chi2$ = 10.26).

Table 4
Class Profiles: 35 Most Representative Words Associated With Class 2 and Class 3

Class 2					Class 3				
Form	*N* in class	Total *N*	% in class	χ^2	Form	*N* in class	Total *N*	% in class	χ^2
Participant	22	29	75.86	43.10	Quantum	32	35	91.43	143.46
Task	22	29	75.86	43.10	Brain	15	15	100.00	73.98
Experience	20	28	71.43	34.67	Consciousness	16	20	80.00	57.21
Target	10	10	100.00	31.15	Conscious	12	15	80.00	42.56
Response	18	26	69.23	29.14	Orch or	8	8	100.00	39.01
Predict	12	14	85.71	28.84	Entanglement	9	10	90.00	37.81
Learn	9	9	100.00	27.99	Moment	9	10	90.00	37.81
Anomalous	10	11	90.91	26.55	System	14	22	63.64	34.52
Person	14	19	73.68	25.46	Retrocausality	7	7	100.00	34.08
Time	21	35	60.00	25.06	Space time	7	7	100.00	34.08
Nonintentional	8	8	100.00	24.84	Microtubule	6	6	100.00	29.16
Spontaneous	8	8	100.00	24.84	Nonlocality	6	6	100.00	29.16
Investigate	10	12	83.33	22.76	Superposition	6	6	100.00	29.16
Stimulus	17	27	62.96	22.40	Reality	7	8	87.50	28.12
Performance	7	7	100.00	21.70	Imply	5	5	100.00	24.26
Precognition	19	33	57.58	20.43	Macroscopic	5	5	100.00	24.26
Subject	11	15	73.33	19.68	Mechanic	5	5	100.00	24.26
Emotional	6	6	100.00	18.57	Paa	5	5	100.00	24.26
Prestimulus	6	6	100.00	18.57	Universe	5	5	100.00	24.26
Mind	13	20	65.00	18.19	Basic	6	7	85.71	23.33
Control	12	18	66.67	17.69	Causality	7	9	77.78	23.53
Feel	7	8	87.50	17.29	Framework	7	9	77.78	23.53
Randomly	7	8	87.50	17.29	Fundamental	7	9	77.78	23.53
Read	7	8	87.50	17.29	Principle	6	7	85.71	23.33
Word	7	8	87.50	17.29	Paper	10	17	58.82	21.26
Reward	8	10	80.00	16.82	Anticipatory	4	4	100.00	19.38
Belief	13	21	61.90	16.31	Biology	4	4	100.00	19.38
Human	12	19	63.16	15.71	Ekt	4	4	100.00	19.38
Go nogo	5	5	100.00	15.45	Paradox	4	4	100.00	19.38
Hit	5	5	100.00	15.45	Nature	10	18	55.56	19.14
Trust	5	5	100.00	15.45	Physical	6	8	75.00	19.01
Twin	5	5	100.00	15.45	Backward	5	6	83.33	18.60
Wander	5	5	100.00	15.45	Bilk	5	6	83.33	18.60
PMIR	5	5	100.00	15.45	Usual	5	6	83.33	18.60
Remote viewing	5	5	100.00	15.45	Causal	8	13	61.64	18.32

Note. *N* = number. Critical values of χ^2 with 1 df: 19.51 for $p < .00001$: 15.14 for $p < .0001$; 10.83 for $p < .001$.

Relationships Between Classes

The relationships of proximity and opposition between the classes may also be observed in the graphic provided by the correspondence analysis performed by the Iramuteq program. As may be seen in Figure 2, the most representative words of the four classes are positioned on two axes: the horizontal axis explains 40.75% of the inertia (i.e., variance) and the vertical axis 30.49%.

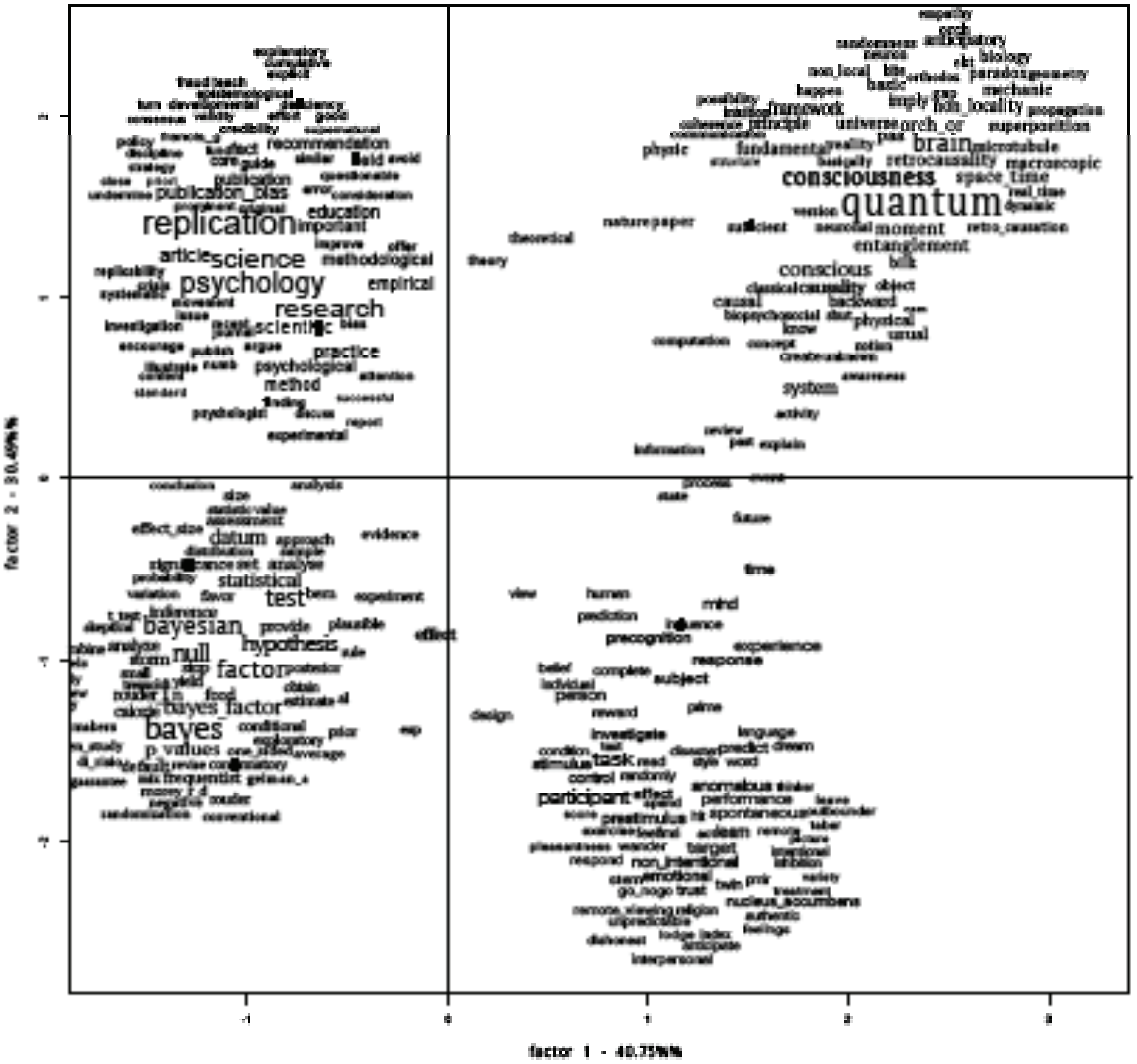


Figure 2. Correspondence analysis

Figure 2 confirms, on the horizontal axis, the strong contrast between Classes 1 and 4 (The Role of Replication in Psychology Research, Bayesian Statistical Inference), and Classes 2 and 3 (Experimental Studies of Anomalous Experiences, Quantum Phenomena and Theories). If we take also into consideration the position of the words on the vertical axis, we may further note that the texts on Replication are the most closely related, in terms of vocabulary, with the texts on Bayesian Statistical Inference, and most distant from the texts on Experimental Studies of Anomalous Experiences. On the other hand, the texts on Quantum Phenomena and Theories are the most closely related, in terms of vocabulary, with the texts on Experimental Studies of Anomalous Experiences and most distant from the texts on Bayesian Statistical Inference.

Discussion

The analysis performed on the summaries of published texts with regard to Bem's controversy extracted four classes, which enabled us to identify four types of impact on the scientific literature on Bem´s (2011) article "Feeling the Future." If Classes 1 and 4 (The Role of Replication in Psychology Research and Bayesian Statistical Inference) might be considered as coming from a more critical point of view, the data also show that they are close to each other and both are related with the identification of methodological problems and questionable research practices, including proposals for solving the problems. Classes 2 and 3 (Experimental Studies of Anomalous Experiences and Quantum Phenomena and Theories), on the other hand, do not share the same point of view, inasmuch as they are characterized by a specific vocabulary which indicates different approaches to Bem´s article.

Data suggest that, subjacent to the Replication class, and to some extent also the Bayesian class, it is possible to consider the question of the authenticity or ontological reality of psi phenomena. Is there evidence or strong evidence that psi is real? The lack of replication of some studies (not only in psi research, but, for example, also in mainstream psychology and medicine), questionable research practices, and the debate around some statistical inference methods, fomented criticisms and negative reactions to Bem´s findings. The Bayesian vocabulary was present more in 2011, when the debate began—for instance, in the same issue of *JPSP* in which the article by Bem appeared, Wagenmakers, Wetzels, Borsboom, and van der Maas (2011) published a paper on Bayesian statistics as a comment on Bem. On the contrary, 2015 was the year in which the question of replication seemed to be most prominent. In that same year, the journal *Science* listed the transformations made by psychologists in their research efforts (i.e., replications of key studies; creation of new models of scholarly publication and peer review, such as preregistration) as one of the notable scientific breakthroughs of 2015 (Bohannon, 2015).

The technical and complex replication topic, and subtopics such as meta-analysis, have been addressed by some parapsychologists over the years (e.g., Honorton & Ferrari, 1989; Watt & Kennedy, 2015, 2016). On the other hand, controversies between the supporters of frequentist and Bayesian forms of interpreting probability have a long history in science, with the former seeing probability as a limiting ratio in a sequence of repeatable events, and the latter as a mental construct that represents uncertainty and which applies not directly to events but to our knowledge about them. As argued by Howey (2002), ". . . a choice between the two interpretations of probability is not forced by pure logic or the mathematics of the situation, but rather depends on the experiences and aims of the individuals involved and their views of the correct form of scientific inquiry" (p. i).

The fact that the Replication and Bayesian classes were represented more in psychology texts confirms our hypothesis that psychology sources have a more critical position concerning Bem´s (2011) article. Firstly, it was published in one of the most influential journals in social psychology, using common methodological paradigms of mainstream psychology, which were reversed; secondly, it is well established that psychologists in general have had a strong critical position about the existence of psi phenomena; thus, it was expected that the most critical reaction would come from psychology sources (although the *Journal of Parapsychology* is also indexed in the subject area Psychology, the relatively small number of articles it published on this topic reduced its impact in the scientific literature indexed in the selected database). When the sources indexed in the subject area Psychology were divided into Only Subject Area Psychology and Psychology and Other Subject Areas, the Bayesian class was represented more in the latter, which is consistent with the usual technical argumentation, particularly in mathematical terms.

Classes 2, Experimental Studies of Anomalous Experiences, and 3, Quantum Phenomena and Theories, underlie two other types of impact of Bem´s (2011) article. As pointed out before, the class Experimental Studies of Anomalous Experiences includes terms from classical empirical research, applied in anomalous experiences studies, and also concepts, methodologies, and theories specific to parapsychological research. The data suggest that, for the authors of these texts, the citation of Bem´s article was applied in a context of process-oriented research on anomalous experiences (experimental work designed to find evidence about the characteristics of processes that underlie anomalous experiences), as

opposed to proof-oriented research, that is, research that aims to find evidence for the existence of psi. This class is represented more in 2012 and in texts whose sources are indexed in subject areas other than psychology, thus confirming the hypothesis that sources from other areas of knowledge might have a more positive position concerning Bem's article.

The Quantum Phenomena and Theories class gives the same result, being more represented in texts whose sources are indexed in other areas of knowledge. In this case, data indicate that quantum theories of brain/consciousness function may leave the door open to the possibility of the existence of psi phenomena. For instance, retrocausality and predictive anticipatory activity (PAA) can be addressed in terms of quantum phenomena. This class, like the Bayesian Statistical Inference class, is represented more in the beginning of the debate (i.e., 2011). Scientists with a favorable position integrated the evidence of retrocausality phenomena as deserving further explanation, whereas the scientists with a more critical position pointed to the necessity for a new method of statistical inference to demonstrate that Bem's (2011) analysis had weaknesses and was not valid. The data trends in the present study suggest that the replication path is currently the most referred to in the attempt to reach closure in this controversy.

References

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407–425.

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000 Research, 4*, 1188. doi: 10.12688/f1000research.7177.1

Benzécri, J.-P. (1973). *L'analyse des données. Volume II. L'Analyse des Correspondances* [Data analysis Volume II. The correspondence analysis]. Paris, France: Dunod.

Bohannon, J. (2015). Reproducibility in psychology. *Science, 350(6267)*, 1459.

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

Chabchoub, A. S. (2008). Une application de la statistique textuelle à l'excès de confiance des dirigeants: Cas des acquisitions.[An application of the textual statistic to managers' excess of confidence: The case of acquisitions]. *9es Journées Internationales d'Analyse Statistique des Données Textuelles*, 1081–1091.

Duggan, M. (2015). *Feeling the future of Bem´s findings*. Retrieved from http://tanclab.org/blog

Gad-el-Hak, M. (2011, January 7). Room for debate: When peer review falters. *New York Times.* Retrieved from http://www.nytimes.com/roomfordebate/2011/01/06/the-esp-study-when-science-goes-psychic/when-peer-review-falters

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

Howie, D. (2002). *Interpreting probability: Controversies and developments in the early Twentieth Century.* Cambridge, UK: Cambridge University Press.

Machamer, P., Pera, M.,& Baltas, A. (Eds.) (2000). *Scientific controversies: Philosophical and historical perspectives*. New York, NY: Oxford University Press.

McClenon, J., Roig, M., Smith, M. D., & Ferrier, G. (2003). The coverage of parapsychology in introductory psychology textbooks: 1990–2002. *Journal of Parapsychology, 67*, 167–179.

McClenon, J. (1982). A survey of elite scientists: Their attitudes toward ESP and parapsychology. *Journal of Parapsychology, 46*, 127–152.

McConnell, R. A., & Clark, T. K. (1991). National Academy of Sciences' opinion on parapsychology. *Journal of the American Society for Psychical Research, 85*, 333–365.

MSNBC (2011). *Professor: Strong evidence ESP is real*. Retrieved from http://www.nbcnews.com/video/msnbc-news/40965642

Palmer, J. (2015). Implicit anomalous cognition. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 215–229). Jefferson, NC: McFarland.

Pélissier, D. (2016). *Initiation à la lexicométrie. Approche pédagogique à partir de l'étude d'un corpus avec le logiciel IraMuTeq* [Initiation to lexicometric analysis. Pedagogical approach using the study of a corpus with the software IraMuTed]. Retrieved from http://www.iramuteq.org/documentation

Poeschl, G., Valentim, J., & Silva, B. (2015). The thousand shades of the financial and economic crisis: A study of the social representations of the crisis in Portugal. *Spanish Journal of Psychology, 18*, 1–10. DOI: http://dx.doi.org/10.1017/sjp. 2015.73.

Ratinaud, P. (2009). IRAMUTEQ*: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires* [IRAMUTEQ: Interface of R for the multidimensional analyses of texts and questionnaires] [computer software]. Retrieved from http://www.Iramuteq.org

Reinert, M. (1983). Une méthode de classification descendante hiérarchique: Application à l'analyse lexicale par contexte [A method of downward hierarchical classification: Application to the lexical analysis by context]. *Cahiers de l'Analyse des Données, 3*, 187–198.

Reinert, M. (1985). Un logiciel d'analyse lexicale: ALCESTE. [A software for lexical analysis: ALCESTE]. *Cahiers de l'Analyse des Données, 4*, 471–484.

Reinert, M. (1986). Classification descendante hiérarchique: Un algorithme pour le traitement des tableaux logiques de grandes dimensions [Downward hierarchical classification: An algorithm for the analysis of logical tables of large dimensions]. In E. Diday et al. (Eds), *Data analysis and informatics* (pp. 23–28). Amsterdam, Netherlands: Elsevier Science.

Reinert, M. (1993). Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars [The "lexical worlds" and their "logic" through the statistical analysis of a corpus made of accounts of nightmares]. *Langage et Société, 66*, 5–39.

Roe, C. (2016). What are psychology students told about the current state of parapsychology? *Mindfield, 7*, 86–91.

Tressoldi, P., Rabeyron, T., Duggan, M., & Bem, D. J. (2016). Feeling the future: A meta-analysis of 82 experiments on the anomalous anticipation of random future events. *F1000Research*. Retrieved from https://f1000research.com/articles/4-1188/v2

Viaud, J., Uribe, J., Acosta T. (2007). Représentations et lieux communs de la mondialisation [Representations and commonly accepted views of globalization]. *Bulletin de Psychologie, 60*, 21–33.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*, 426–432.

Watt, C., & Kennedy, J. (2015). Lessons from the first two years of operating a study registry. *Frontiers in Psychology, 7.* DOI: http://dx.doi.org/10.3389/fpsyg.2015.00173

Watt, C., & Kennedy, J. (2016). How prospective meta-analysis could be useful for parapsychology. *Abstracts of Presented Papers: The First Combined Convention of the Parapsychological Association (59th) and the Society for Scientific Exploration (35th),* p. 49.

Zingrone, N. L. (2006). *From text to self: The interplay of criticism and response in the history of parapsychology* (Doctoral dissertation, University of Edinburgh). Retrieved from https://www.era.lib.ed.ac.uk/handle/1842/7722

Center for Psychology at University of Porto
University of Porto
Porto, Portugal
bruno_silva@fpce.up.pt

Abstracts in Other Languages

French

L'ARTICLE "PRESSENTIR LE FUTUR" DE BEM (2011) CINQ ANS PLUS TARD : SON IMPACT SUR LA LITTERATURE SCIENTIFIQUE

RESUME : Cette étude analyse l'impact sur la littérature scientifique de l'article controversé de Bem (2011), "Pressentir le futur". Les textes qui citent l'article de Bem ($N = 162$) furent identifiés par la base de données Scopus d'Elsevier pour les années 2011 à 2015. Leurs résumés furent analysés par le programme Iramuteq pour les données textuelles. L'analyse suggère que l'impact peut être regroupé en quatre classes : (a) La classe de Réplication est caractérisée par un vocabulaire tourné vers la question du rôle de la réplication dans la recherche en psychologie ; (b) La classe Bayésienne réflète les mérites apparents de cette approche par rapport aux statistiques inférentielles plus traditionnelles, c'est-à-dire les statistiques s'appuyant sur les valeurs p ; (c) la classe des Etudes expérimentales des expériences anomales (AE) inclue des termes relatifs à la recherche empirique classique, appliquée aux études d'AE, et les concepts, méthodologies et théories

propres à la recherche parapsychologie ; (d) la classe des Phénomènes et théoriques quantiques suggère que des théories quantiques du cerveau et de la conscience pourraient ouvrir la porte à la possibilité de l'existence des phénomènes psi. Les classes de Réplication et Bayésienne sont les plus représentées dans la littérature psychologique, confirmant notre hypothèse que les sources psychologiques sont davantage dans une position critique. Les données suggèrent également que la réplication est la voie à laquelle le plus grand nombre se réfère dans une tentative pour clore la controverse.

German

BEM'S "FÜHLEN DER ZUKUNFT" FÜNF JAHRE DANACH: IHR EINFLUSS AUF DIE WISSENSCHAFTLICHE LITERATUR

ZUSAMMENFASSUNG: Die Studie untersucht den Einfluss des 2011 erschienenen umstrittenen Artikels von Bem: „Feeling the Future" auf die wissenschaftliche Literatur. In der Elsevier Scopus-Datenbank von 2011 bis 2015 wurden $N = 162$ Texte identifiziert, die Bems Artikel zitieren. Ihre Abstracts wurden unter Verwendung des Iramuteq-Programms für Textdaten analysiert. Die Analyse ergab, dass ihr Einfluss in vier Klassen eingeteilt werden kann: (a) In der Replikationsklasse dominiert ein Sprachgebrauch, der die Rolle der Replikation in der psychologischen Forschung beschreibt; (b) die Bayesklasse hebt die Vorteile dieses Zugangs hervor, wenn er mit den mehr herkömmlichen schließenden Statistiken verglichen wird, und dies besonders bei Statistiken, die auf *p*-Werten beruhen; (c) die Klasse der Experimentellen Studien zu Anomalen Erfahrungen (AE) umfasst Begriffe, die sich auf klassische empirische Forschungen beziehen, angewandt auf AE-Studien und Konzepte, Methodologien und Theorien, die charakteristisch für die parapsychologische Forschung sind; (d) die Klasse der Quantenphänomene und Theorien legt nahe, dass Quantentheorien über die Funktionsweise von Gehirn/Bewusstsein die Möglichkeit offenlassen, dass Psi-Phänomene tatsächlich existieren. Die zur Replikation und zu Bayes gehörenden Klassen finden sich eher in der psychologischen Literatur und bestätigen unsere Hypothese, dass psychologische Quellen eine eher kritische Position einnehmen. Darüberhinaus legen die Daten nahe, dass auf Replikation gegenwärtig am häufigsten Bezug genommen wird beim Versuch, die Kontroverse zu beenden.

Spanish

EL ARTÍCULO DE BEM "FEELING THE FUTURE" (2011) CINCO AÑOS DESPUÉS: SU IMPACTO EN LA LITERATURA CIENTÍFICA

RESUMEN: Este estudio analiza el impacto en la literatura científica del controvertido artículo de Bem "Feeling the Future" de 2011. Los textos que citan el artículo de Bem ($N = 162$) fueron identificados en la base de datos de Elsevier Scopus para los años 2011 a 2015. Sus resúmenes fueron analizados utilizando el programa Iramuteq para datos textuales. El análisis sugiere que el impacto se puede agrupar en cuatro clases: a) La clase de Replicación se caracteriza por un vocabulario que trata del papel de la replicación en la investigación psicológica; b) la clase Bayesiana refleja los méritos percibidos de este enfoque cuando se compara con estadísticas inferenciales más tradicionales, es decir estadísticas basadas en los valores de *p*; c) la clase de Estudios Experimentales de Experiencias Anómalas (EA) incluye términos relacionados con la investigación empírica clásica, aplicada a estudios de EA, y conceptos, metodologías, y teorías específicas a la investigación parapsicológica; (d) la clase de Fenómenos y Teorías Cuánticos sugiere que las teorías cuánticas de la función cerebro/conciencia pueden dejar la puerta abierta a la posibilidad de la existencia de fenómenos psi. Las clases de Replicación y Bayesiana están representadas más en la literatura psicológica, confirmando nuestra hipótesis de que las fuentes en psicología tienen una posición más crítica. Los datos sugieren además que la replicación es actualmente la vía más referida para tratar de llegar a una conclusión en esta controversia.

Journal of Parapsychology, 81(1), 46-62

ANOMALOUS COGNITION, DISSOCIATION, AND MOTOR AUTOMATISMS[1]

By John Palmer

ABSTRACT: 80 volunteers completed an anomalous cognition task reflecting motor automatism, preceded by a progressive relaxation exercise. Pasted on a computer writing tablet was a 4x4 grid of 16 1-inch squares each containing a number 1–4. For 36 trials, participants (Ps) explored the grid with the computer pen, recording their responses to the randomly selected target squares by stopping for 1 s. They were assigned to 4 cells in a 2x2 design with the independent variables being hand used (right or left) and a dissociation facilitator (mind-blanking with eyes closed vs. reading quotations on a screen). The dependent variable was location hits, an unweighted combination of standardized square and quadrant hits. ANOVA revealed significant psi-missing in the quotations/left condition and significant psi-hitting elsewhere (EQR). High scores on the Detachment (DET) component of the Dissociative Processes Scale (DPS) and reports of the hand being moved by an outside force (OF) during the AC task jointly predicted high location hit scores in the EQR condition, as did DPS Imagination (IMA). IMA and DET correlated significantly with hits on number across all conditions. This exploratory study was interpreted as reflecting psi mediation by a motor process and a cognition process on different trials.

Keywords: anomalous cognition, dissociation, motor automatism, state-trait

The general hypothesis tested in the experiment is that psi is facilitated by dissociated states of consciousness. Underlying this hypothesis is the more basic premise that the relative absence of overt manifestations of psi comes not from the lack of psi information in the unconscious mind, but rather by awareness of this information being blocked by psychological defense mechanisms. One vehicle by which these defense mechanisms might be circumvented is a dissociated state of consciousness, as exemplified by Ernest Hilgard's (1977) "hidden observer," which under hypnosis can recall mentation that is inaccessible to normal waking consciousness.

The application of the concept of dissociation to psychic phenomena can be traced back to F. W. H. Myers (1903). Even though he did not use the term much himself, dissociation is clearly what he had in mind when he introduced the notion of the secondary self, which can function independently of the primary personality or supraliminal consciousness. The secondary self often manifests through what Myers called sensory or motor "automatisms." Myers included under motor automatisms automatic writing, automatic speaking, automatic drawing, and use of the pendulum. I consider motor automatisms to represent a purer form of dissociation than sensory automatisms, because motor automatisms in their unadulterated form do not involve any conscious cognition.

Seeking an anomalous cognition (AC) test related to motor automatisms that participants (Ps) could relate to easily, I decided to employ the Ouija board concept (Palmer, 2011). However, because use of the Ouija board for conjuring spirits has been shown to have adverse psychological effects in some cases (e.g., Palmer, 2001), I decided to make the association to the Ouija board somewhat more distant. As part of this effort, I substituted an alphabet board consisting of the 26 letters arrayed on the four outer sides of the board. The target words were one-syllable homographs. Ps were instructed to repeatedly move a pointer or planchette randomly around the board until they felt the impulse to stop on a particular letter, at which time they recorded the letter on a notepad. In the meantime, a sender in another room on a different floor of the building was asked to transmit the target.

Prior to the session, participants filled out a trait measure of dissociation, consisting of Persinger's Complex Partial Epileptic Signs scale (Persinger & Makarec, 1993), with scores on the Tellegen Absorption

[1] As I am Editor of the *JP*, to avoid self-serving bias I followed my policy in such cases and had this paper peer-reviewed under the auspices of Dr. Richard Broughton of the University of Northampton.

Scale (Tellegen, 1978) partialled out. At the end of the session, Ps were shown the five words in the target set and asked to rate each word on a 0–20 scale based on its correspondence to letters they had gotten from the board, as well as mental imagery they had spontaneously experienced at any point during the session. They also completed a rating scale on their experiences during the session, the key item of which was what percentage of the movements of their hand they felt had been guided by an outside force (OF), a state measure of dissociation. To allow for assessment of a linear trend, the sample was divided into three roughly equal groups: none of the time, 1-40% of the time, and 40-100% of the time. The 11 of 40 Ps who reported that their hand was so guided 1–40% of the time demonstrated a strong AC effect, whereas Ps in the other groups averaged close to chance. I speculated that the subjective experiences of dissociation in the 1–40% group tended to differ qualitatively from those in the 41–100% group, but I did not interview the Ps about their experiences. The trait dissociation measure correlated significantly positive with the OF question but not with the AC scores.

The present experiment was an extension of Palmer (2011) with the following specific objectives: (a) to attempt to replicate the OF/AC correlation and to understand it better by interviewing those Ps who answer the OF question positively about the precise nature of their experience of the outside force; (b) to assess a possible relationship between trait dissociation and AC using a more direct measure of dissociation; (c) to explore the effects of additional procedures to facilitate state dissociation during the task; and (d) to explore possible differences in scoring as a function of whether Ps use the right or left hand for their AC responses. The experiment also differed methodologically from Palmer (2011) in that the task was forced choice rather than free response, and there was no sender.

Watson's (2003) Dissociative Processes Scale (DPS) was chosen as the measure of trait dissociation. One reason is that it is more discriminative within the normal range than the standard but highly clinical Dissociative Experiences Scale. Second, it contains items that seemed to reflect the particular kind of dissociative process I was interested in, which can be represented by the metaphor of one part of one's mind keeping a car on the road while the other part is engaged in conversation with the person in the passenger seat. The DPS has three factors or subscales: Obliviousness, Imagination, and Detachment. The items that struck me as most relevant to the present experiment are in the subscale measuring Obliviousness, which Watson (2003) defines as "the tendency to engage in mindless and automatic behaviors and to enter into naturally occurring trance states" (p. 300). Thus, I based my hypothesis on the Obliviousness subscale.

Although parapsychological studies comparing motor performance with the right and left hand have led to rather ambiguous results (Broughton, 1976; Maher & Schmeidler, 1977), left-hand performance was found to be superior in a psychology experiment in which Ps had to match by touch the directional orientation of stainless steel rods they could not see to other rods with diverse orientations (Benton, Varney, & Hamsher, 1978). On this basis, along with the general tendency for AC to be associated with right-hemisphere processing (Braud, 1975), I gingerly predicted superior performance with the left hand.

The major hypotheses were as follows:

Hypothesis 1: Performance on an AC task providing for motor automatisms will be better for Ps who have the experience that the associated hand is being moved by an OF during the task than for other Ps.

Hypothesis 2: Performance on an AC task providing for motor automatisms will be positively related to Ps' tendencies toward dissociation as measured by the Obliviousness subscale of the Dissociative Processes Scale.

Hypothesis 3: Among right-handed Ps, performance on an AC task providing for motor automatisms will be better if the AC responses are made with the left hand than with the right hand.

Hypothesis 4: Dissociative tendencies as measured by the Obliviousness subscale of the DPS will be greater in Ps who feel their hand was being moved by an OF during the AC task than in other Ps.

Method

The study was approved in advance by the Rhine Research Center (RRC) Institutional Review Board.

Participants

Eighty mentally healthy adult volunteers were recruited to serve as Ps. An announcement was sent out to the RRC mailing list of more than 3,000 persons who had contacted the RRC in the past, but almost all the respondents were from the local area. The letter indicated that I was especially interested in testing people who found the rationale behind the experiment congenial. Ps were contacted, usually by email but occasionally by phone, in the order I received their responses. Once a session time had been established, I emailed them the DPS as well as the RRC's Participant Information Form, which they either returned by email or brought with them to the session.

Main Measures

Dissociative Processes Scale (DPS). The DPS consists of 33 items (Watson, 2003). Responses are made on a 5-point scale from *strongly agree* to *strongly disagree*. Cronbach's alpha was .93 at original testing and .94 at retest. The scale consists of three factorially derived subscales: Obliviousness (14 items), Detachment (6 items), and Imagination (7 items). The other six items are presumably buffers. Cronbach alphas for the scales range from .85 to .89.

Post-experiment questionnaire. Following the test session, I (the experimenter) asked the Ps a series of questions concerning their experience during the session and recorded their responses on a rating scale. The questions were presented in a semistructured interview format, which allowed me to better understand Ps' experiences and to detect any misunderstandings of the questions by providing for follow-up questions, such as requests for concrete examples of relevant experiential elements. The most important question was taken directly from Palmer (2011): "Did you have the impression at any time during the session that the pen was being guided by an outside force? (*yes*; *no*). If yes, about what percentage of the movements were caused in this manner? (*1–20%*; *21–40%*; *41–60%*; *61–80%*; *81–100%*)." Other questions concerned mainly Ps' expectations of success in the task and how well they were able to implement the task procedure.

The AC Task

Apparatus. The AC task used a Dell GX240 desktop computer and monitor, a computer writing tablet (Addesso Cyber Tablet 1200), and a pen that came with the tablet. The basic idea is that Ps move the pen over the surface of the tablet and stop at a particular location to register their "guess."

Target board. The tablet is 13.5" x 16.5" and has a writing area of 9" x 12." The writing area is covered by a sheet of flexible transparent plastic attached to the back side of the writing area. A 4" square piece of writing paper was taped under the plastic cover so as to cover the middle of the writing area. This target area is referred to hereafter as the grid. Three strips of soft wood (5 cm high) were attached to the outside of the plastic cover along the right, left, and back sides of the grid. The purpose was to keep P from moving the pen outside the target area. For the same purpose, a 1 mm thick piece of picture-hanging tape was placed along the front side of target area. It was not as steep as the other barriers, to allow P to reach the bottom of the grid with the pen without making it uncomfortably vertical. The tablet and grid are illustrated in Figure 1.

The grid was formatted as a 4 x 4 square matrix consisting of 16 smaller squares, each 1" x 1". The matrix was divided into four quadrants, with each quadrant consisting of four squares. In each square was printed a single digit from 1 to 4, such that within each quadrant each number appeared once and only once. The numbers were distributed in a quasi-random fashion such that they were balanced with respect to their locations in the matrix.

For each trial, P explores the grid by moving the pen over the surface of the cover, in contact with it at all times. P makes a response by stopping the pen at the chosen location for 1 s. The identity of the square under the tip of the pen is then sent to the computer as P's response for the trial.

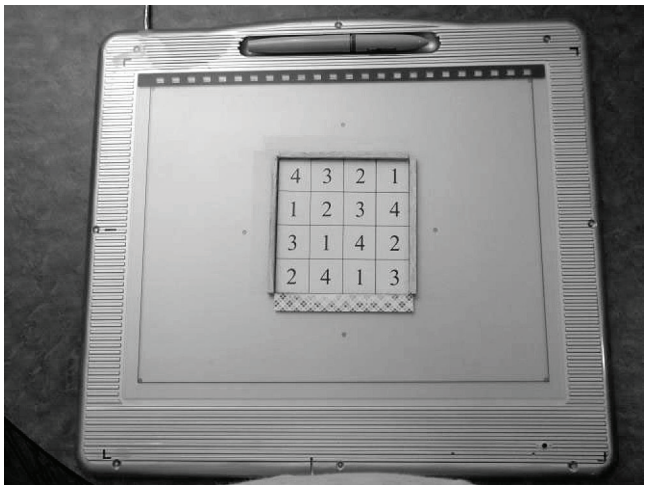


Figure 1. Target board with pen

Target sequences. The targets were multiple-aspect, in that each could identified by square (P = 1/16), quadrant (P = 1/4), and number (P = 1/4).

At the beginning of the experiment, John Kruth, Executive Director of the RRC, generated 80 successive random sequences of 36 square targets (1 to 16). To do so, he used an algorithm developed by Marsaglia and Zaman (1987), which has been thoroughly tested for randomness. To avoid intuitive decisions, Kruth was instructed to use 1 and 2 as the seed numbers. This sequence was emailed to the programmer, Daniel Ruszaj, who incorporated it as a file in the software package he was developing to create the AC scores. This program, written in Java, compared the targets with P's responses to calculate how many square, quadrant, and number hits each P obtained. For example, if the target was the square in the upper left corner, and P chose the square next to it on the right, P would obtain a quadrant hit but not a square or number hit. This procedure allowed the experimenter to remain blind to the targets throughout the experiment.

Attempts to Facilitate a Dissociative State

Task manipulation. The purpose was to disengage the conscious mind from the AC task.

Eyes-closed. In this condition, Ps were to blank the mind with eyes closed and wait for an impulse to stop the hand movement and make a response. It was emphasized that this impulse should be a feeling in the hand, not a mental or cognitive image.

Quotations. In this condition, Ps were asked to distract the conscious mind by reading a series of quotations on the computer screen.

A pool of 100 quotations was created from *Bartlett's Familiar Quotations* (Bartlett, 2002). In the book, the quotations are arranged by the birth year of the author of the quote. I started at 1900 and worked forward. The list was long enough so that no P would read any quote twice. A separate list of six quotations was created for the practice sessions. These quotes did recycle. The long list was randomized before testing began.

The 19th P indicated during the post-task interview a marked dissociative feeling related to her hand movements after reading quotes she found humorous. This struck me, as the participant in one of the early 20th century automatic writing experiments mentioned that she felt the text she was reading as a distraction task had to be entertaining to be effective (Solomons & Stein, 1896). This P's experience changed my mind about the best quotes to use to achieve the experimental objective. Because I was not keeping track of the AC scores, one reason being that I would not be biased by them if I wanted to make a methodological adjustment after the experiment began, and because I considered the experiment exploratory, I decided to immediately create a new pool of quotes taken from the *Dictionary of Modern Humorous Quo-*

tations (Metcalf, 2009). I chose the ones that were most humorous to me. The same adjustments for length were made as with the original quotes, but overall the new quotes were shorter than the original ones. To compensate, I increased the number of quotes in the task list to 144. Some of them are sexual in nature (but not pornographic; they resemble jokes one finds on late-night television talk shows in the U.S.). Ps were given the opportunity to view instead the original quotes upon indication that they might be offended by the humorous ones. No one requested the original quotes. A separate list of humorous quotes was also created for the practice trials.

The quotes were presented on the computer screen one at a time. As it would be distracting to require P to press a button to make the next quote appear, I had the computer replace each quote with the next quote after a fixed number of seconds. The intent was for each quote to disappear as soon as P had finished reading it. I took two steps to achieve this. First, I tried to make all the quotes of approximately the same length. To accomplish this, I would often choose only some of the sentences from multi-sentence quotations and occasionally omit trivial words within sentences. Second, before the task, I had Ps time themselves reading a typical quote (not in the task list) using a stopwatch. I then set the duration of each quote to match that number. Ps were given the option to change this number after the practice session. Ps were told that if they finished a quote before it changed, they should begin rereading it. If, on the other hand, the quote changed before they finished reading it, they should not worry about it but just move on to the next quote.

Whenever an AC response was registered, the screen briefly flashed. This single flash was intended to assure P that a response had registered, without distracting P from reading the quotes.

During the instruction period, Ps were told that immediately following the AC task they would take a short quiz to measure how well they remembered the quotes they saw. The quiz was a recognition task in which they were shown 10 quotes that had been presented during the AC task and 10 unseen control quotes, all in random order. Ps indicated by a mouse click whether they recognized the quotation from the AC task. It was given solely to motivate the Ps to pay attention to the quotes during the AC task. All Ps got a very high score on it.

Progressive relaxation. The second method used to facilitate a dissociated state of consciousness during the AC task was to induce a mild altered state of consciousness immediately beforehand. The induction consisted of an abbreviated form of the Jacobson progressive relaxation technique, frequently given at the beginning of AC ganzfeld tests. Progressive relaxation is particularly appropriate for a task that provides for motor automatisms, because it includes commands to alternatively tense and relax various muscle groups.

The relaxation suggestions were followed by a review of the instructions for the AC task and suggestions that P would succeed in the task. The entire exercise was recorded on tape by the experimenter and played over speakers to the Ps.

Design

The experimental design was a 2 x 2 factorial with Method (eyes-closed vs. quotations) and Hand (left vs. right) as the independent variables. Hand refers to whether Ps were asked to hold the pen with the right or left hand during the AC task. Twenty Ps were assigned to each of the four cells, abbreviated ER (eyes-closed, right hand), EL (eyes-closed, left hand), QR (quotations, right hand), and QL (quotations, left hand). Condition assignments were made by Ruszaj, who performed a random permutation of the numbers 1–80, with 20 of the numbers coded as representing each of the four conditions. The list was then sent to me so that I could be informed immediately before a session which condition a given P should be in.

Lab Room Layout

P was seated in front of a table in a padded chair that could be tilted backwards and adjusted for height. In the left-hand conditions, the tablet was located to the left of the computer monitor, whereas in the right hand conditions these positions were reversed. Two small audio speakers were located at the back of

the table facing P. Before the AC test, P was given the choice to have the room illuminated by the overhead lights or a small table lamp with a 40W bulb located at the far right end of the table.

Procedure

After collecting any questionnaires the Ps brought with them, I gave them a more detailed description of the procedure than was provided in the solicitation letter. I illustrated how to take the AC task by moving the pen around the grid and stopping for 1 s. In an effort to minimize response biases, I emphasized that they should explore the entire grid before making a selection. I also described the three methods of scoring (square, quadrant, and number) and told them that they would receive feedback of their scores at the end of the phase. Finally, I described the task manipulation (quotations or eyes-closed).

Next, I set up the practice session on the computer. I had Ps adjust the location and orientation of the tablet as well as the height of the adjustable chair for maximum comfort and ease in moving the pen around the grid. Ps in the quotation conditions then timed their quotation reading speed as described above. The practice session had two stages. In the first stage, I began by having Ps look at the grid while moving the pen around. When I was comfortable they could do this easily, I asked them to continue with their eyes closed for a few more trials. To get them adjusted to how long they needed to stop the pen for a response to register, a beep sounded after the pen was motionless for 1 s. In the second practice stage, Ps practiced the procedure to be used in the formal experiment, which did not include the beeping sound. I asked them to continue practicing until they and I agreed that they could do the movements correctly and effortlessly. This practice was particularly important in the quotation conditions, where Ps had to learn to make their AC responses independent of where they were in the process of reading the quotations; the common tendency at the outset is to make the AC response at the time one finishes reading a quotation. Ps varied widely in how long they practiced, and the number of practice trials usually exceeded the 36 trials in the formal experiment.

After the practice was completed, I set up the main experiment, asking Ps in the quotation conditions if they wanted to change the length of time each quotation appeared on the screen. I told Ps in the eyes-closed conditions that a beep would inform them when the 36 trials were done but that they would not hear a sound after each AC response. During the task I was seated in the adjacent room doing professional reading. I closed the door between this room and the test room as I left. Ps were instructed at the end of the AC task (eyes-closed conditions) or the quotation recognition quiz (quotation conditions) to knock on the door to retrieve me.

When I returned, I immediately had Ps fill out the post-task questionnaire, and I discussed their answers with them afterwards. I focused especially on their subjective experience while taking the AC task. I then described the AC feedback screen, which consisted of trial-by-trial scores, total scores, and number of hits needed for a statistically significant AC score for all three scoring modalities (square, quadrant, and number). I told them that because I did not want to see the scores myself, after I left the room they could see their AC score (and quiz score, if applicable) by pressing the <enter> key on the computer keyboard. They should then press the <enter> key again to clear the screen. A paper tablet was available on which they could write down their AC scores; they could take the sheet home with them if they wished. After asking if they had any further questions about the experiment, I paid them $15 in cash and had them sign a receipt. Finally, I thanked them for their participation and left the room.

Pilot Testing

Eight pilot sessions were conducted with six individuals to determine if it was necessary or desirable to make refinements in the procedures and materials as described above. The Ps were drawn from people closely associated with the RRC, and none were included in the formal experiment.

Data Analysis

Although there were three separate AC scores, to obviate multiple analysis issues I decided to

compute a single AC *z*-score to test the hypotheses. I wanted a composite score but I wanted it to reflect the application of a motor automatism as opposed to a strictly cognitive mechanism. For that reason, I omitted the number scores from the conglomerate. Specifically, for each run, the two remaining component scores (quadrant and square) were combined to produce a single *z*-score, the computation of which is described in Results.

The statistical tests used to test the hypotheses included ANOVAs, *t* tests, chi-squares, and Pearson correlations. The need for nonparametric alternatives for the parametric tests was eliminated by applying normalization procedures to highly skewed distributions. The statistical analyses were performed using SPSS software.

Left-handed Ps were removed from analyses involving the Hand variable. I was not interested in hemisphere dominance, just which hand the P customarily uses or is comfortable using. Ambidextrous Ps were retained.

All *p* values are two-tailed expect for tests of the outside-force-AC relationship, which was considered a replication of Palmer (2011)

Results

AC Scores

Performance errors are one of the hazards of asking Ps to perform a task over which they are supposed to have minimal conscious control. In this case, the concern was mechanical errors in moving the pen around the tablet. To minimize these errors is why Ps were given extensive practice before the official task. However, I was never convinced this learning would completely carry over to the formal session. One reason is that Ps were now trying to use psi to obtain a good score. The second reason is that they were presumably in a different state of consciousness as a result of the intervening relaxation exercise.

As I did not observe Ps performing the formal task, I could only get information about possible errors by looking at the data, which I did not do until data collection had been completed. Useful information would be reaction times (the time from the end of trial t to beginning of trial t+1), so I had the computer record this information in milliseconds.

After the experiment, I printed out a sheet for each P with two columns of data. The first column gave the 36 square responses for the run and the second column gave the corresponding reaction times. I noticed that for many Ps some of the reaction times were quite low, as little as 750 ms. In most of these cases the response square was the same for this trial and the preceding trial; in almost all other cases the two response squares were adjacent. During the practice sessions, Ps were encouraged to "explore the grid" before each response by moving the pen randomly around the tablet, so I expected the reaction times to be at least several seconds. On the other hand, my feedback during debriefing suggested that during the actual task Ps would often go directly from one response to the next, which could be done quite quickly. But if that were going on, I would expect the response squares for the two trials to be nonadjacent much more frequently than they in fact were.

Instead, I concluded that most of these responses were inadvertent jiggles of the pen in the area of the first response (despite my efforts to warn Ps to keep the pen still between trials). This would cause an unintended response to register. I had noticed this behavior during the early stages of the practice sessions, but by the time the session ended it had vanished. However, as noted above, there was reason to fear that it might re-emerge during the formal task.

In the quotation conditions, the screen flashed after each response, so Ps knew, consciously or unconsciously, that a trial registered and they could use this information, consciously or unconsciously, to correctly identify the proper target for the next trial by jumping one target ahead. However, this information was not available to Ps in the eyes-closed conditions. I had considered continuing the beeps after each response in the eyes-closed conditions, which would have functioned like the flashes in the quotation conditions. I decided against this, partly because I was afraid the beeps would bring them out of the meditative

state that I expected to be optimized in the eyes-closed conditions, but also because Ruszaj was unable to get the software to produce the beeps reliably.

I decided to make up a set of alternate response sequences that would take this bias into account. Specifically, I wanted to capture the response sequences the Ps intended, not necessarily the ones that were sent to the computer. To do so, I perused each sheet individually to detect trials on which the bias was likely to have occurred. Because the sheets did not include the target sequences, I could do this with no sensory information about hits, and thus I was not biased. I assumed that the quotation Ps adjusted for the bias and the eyes-closed Ps did not. Thus, for quotation Ps, I simply removed the response, meaning that the suspect trial was declared missing. For eyes-closed Ps, however, I needed to adopt a more complicated strategy. Starting at the top of the run, for each suspect trial, I removed the response for that trial and moved all the remaining responses up one space; for example, if Trial 6 was suspect, the response to Target 6 would now become the response to Target 5, the response to Target 7 would become the response to Target 6, and so forth. This left a number of missing (unscored) trials at the end of the run corresponding to the number of suspect trials, as the sequence of *registered* responses had been prematurely exhausted. This exercise resulted in marked changes in the AC scores in the eyes-closed conditions but trivial changes in the quotation conditions.

I also decided to remove from subsequent analyses the run scores of 9 Ps who made a large number of such errors. All these Ps were in the eyes-closed condition; 5 used the right hand and 4 used the left. I did this because I felt the chances were high that at least one of my guesses about whether a response was erroneous would be wrong, rendering the resulting sequence especially unlikely to be a valid representation of the P's intent. (I fully expected that some of my guesses were incorrect even in the other cases, but I believed that overall the revised sequences would be closer than the original sequences to Ps' intended responses.) I also questioned whether the removed Ps were performing the task competently overall. Changes were made to the response sequences of 40 of the remaining 71 Ps (56%).

To create for each P hit totals adjusted for the number of trials, I applied the formula $Y = X*36/N$, where Y is the adjusted score, X is the original score, and N is the number of valid trials. I then converted all scores to z-scores using the formula $(X\text{-}M)/SD$, where X is the number of hits, and M and SD are the mean and standard deviation of the distribution of hits for all valid Ps. These z-scores were obtained for the square, quadrant, and number hits as recorded by the computer for all trials of all Ps (Model 1) and for the corresponding hits obtained using the revised response sequences as described above (Model 2). The hit scores used to test the predictions were an unweighted combination of square and quadrant scores using the Stouffer formula $(z_s+z_q)/\sqrt{2}$, where z_s is the z-score for squares, z_q is the z-score for quadrants. Hereafter, these scores will be referred to as "location zs." They were predesignated as the main dependent variable and the variable used to test the hypotheses.

Normalization. All the hit distributions with the exception of number hits were positively skewed due to the presence of outliers on the right tail. I am strongly opposed to removing outliers just because they are outliers, as doing so grossly violates the integrity of the distribution. To remove the excess skewness, I developed a procedure for systematically moving the outliers toward the mean. Specifically, I find the largest distance between any two adjacent outliers and move this value and all more extreme values on that tail half this distance toward the mean. I repeat this process until the value of the skewness statistic is less than 2.0. This method was applied to all the skewed distributions and the resulting distributions were used for all analyses.

Comparison of models. Model 1 is represented by use of the original 36-trial scores of all 80 Ps. This model is preferable if either of two assumptions is valid. The first assumption is that even in the eyes-closed conditions, Ps were not set off course by their invalid responses, because either (a) they knew, perhaps by psi, that their last registered response was not intended and adjusted their responses accordingly to match the real targets for all future trials, or (b) through some less mechanical means they simply "made things turn out right." The latter has been referred to as psi being "goal-oriented." Often cited as evidence for goal-orientation is a random event generator (REG) experiment in which results were equally strong regardless of the internal complexity of the REG (Schmidt & Pantas, 1972). Model 1 is also valid if the observational theories (Millar, 1978) were operative, because the Model 1 scores were the ones shown to

the Ps at the end of the session. Model 2 is preferable if Ps in the eyes-closed condition were not aware that they had registered an unintended response and attempted to match their response to the target of the preceding trial.

I decided to choose between the models empirically. The first test one normally does in evaluating data is a global analysis (e.g., an ANOVA) that registers the effects of the manipulated independent variables, in this case, Method (eyes-closed vs. quotations) and Hand (right vs. left), on the dependent variable, location *z*s. I decided to do the same ANOVA for Model 1 and Model 2 and compare the results. If the total *F* for one or both models is significant, I would choose for subsequent analyses the model that is most significant (highest R^2). If neither model is significant, I would choose the model that yields the highest magnitude mean *z*-score (hitting or missing direction) across all Ps in the model. If there is no psi in the data, it doesn't matter which model I choose; all I will get from either one are Type-I errors, and there is no reason to expect more of these from one than from the other. If there is psi in the data, then the best guess is that the model that shows the stronger results at this stage is the more valid. The only hypothesis tested by these ANOVAs is the handedness hypothesis (Hypothesis 3). All the other hypotheses involve individual-difference measures as predictors and are not dependent on the ANOVA. These were to be tested using only one model, thus eliminating the multiple-analysis problem insofar as it relates strictly to the dependent variable.

The ANOVAs represented a 2 x 2 factorial, between-subjects design, with Method (eyes-closed vs. quotations) and Hand (right vs. left) as the independent variables. The four cells are ER (eyes-closed, right hand), EL (eyes-closed, left hand), QR (quotations, right hand), and QL (quotations, left hand). Location *z* was the dependent variable. Because one of the independent variables represented which hand was used for the task, the five left-handed Ps were removed from the ANOVAs.

The means and standard deviations for location *z* in each cell, left-handers removed, are reported in Table 1 for both Model 1 ($N = 75$) and Model 2 ($N = 68$). Only seven disqualifiers were removed, because two were left-handed and removed on that basis. The Model 1 ANOVA was significant overall, $F(3,71) = 3.56$, $p = .018$ ($R^2 = .13$). The only significant effect was the Method x Hand interaction, $F(1,71) = 6.66$, $p = .012$. Inspection of the means reveals that the interaction consisted of positive scoring (psi hitting) in the EL and QR conditions, offset by negative scoring (psi missing) in the ER and especially the QL conditions.

Model 2 was also significant overall, $F(3,64) = 5.34$, $p = .002$ ($R^2 = .20$). Significant main effects for Method, $F(1,64) = 5.38$, $p = .024$, and Hand, $F(1,64) = 5.08$, $p = .028$, were superseded by a significant Method x Hand interaction, $F(3,64) = 3.56$, $p = .018$. As with Model 1, the interaction reflected positive scoring in the ER, EL, and QR conditions offset by negative scoring in the QL condition. Because R^2 was higher in Model 2 than in Model 1, Model 2 was chosen for the subsequent analyses. Model 1 is not discussed further and in fact was never analyzed.

For subsequent analyses of Model 2, conditions ER, EL, and QR were pooled to form condition EQR. The reason for isolating the QL condition is as follows. In an earlier paper (Palmer, 1975), I argued that when an AC mean is below chance, one should expect a negative relationship between AC and a predictor, and this is what I found in a series of OBE/ganzfeld experiments I conducted many years ago (e.g., Palmer, Bogart, Jones, & Tart, 1977). The rationale for this prediction is that the Ps with the lowest scores are producing the most psi, with Ps scoring low on the predictor averaging around chance. Thus, as I had hypothesized positive relationships between AC scores and dissociation, it is at least highly questionable to apply that prediction to the QL condition, where there was significant overall psi-missing. Three left-handers were also removed from the EQR condition, reducing the total *N* from 71 to 68. The mean number of location hits for the 49 Ps in the EQR condition was 0.42 ($SD = 1.12$), $t(48) = 2.62$, $p = .012$. The mean for the 19 Ps in the QL condition was -0.76 ($SD = 0.94$), $t(18) = 3.53$, $p = .002$.

It is obvious from these results that Hypothesis 3, better performance with the left hand, was not supported.

Number *z*. Because number hits are independent of location hits, the cell means were added to Table 1. The overall model for number *z*s is nonsignificant, $F(3,64) = 0.58$, $p = .63$. Neither the main effects nor the interaction were significant. Because the QL mean was the lowest of the four AC cell means, I compared the EQR and QL conditions on number hits. The result is nonsignificant, $t(66) = 0.92$.

Table 1
Mean (n) for Model 1 (N =75) Location, and Model 2 (N = 68) Location and Number, z scores as a Function of Method and Hand

	Location *z* (Model 1)		Location *z* (Model 2)		Number *z* (Model 2)	
	E	Q	E	Q	E	Q
R	-0.07 (18)	0.38 (19)	0.46 (14)	0.39 (19)	-0.01 (14)	0.21 (19)
L	0.14 (19)	-0.75 (19)	0.41 (16)	-0.76 (19)	-0.11 (16)	-0.21 (19)

Note. E = Eyes-closed. Q = Quotations. R = Right hand. L = Left hand. Left-handers were removed.

State Dissociation (Outside Force) and AC

Change in hypothesis. The state measure of dissociation was the question "Did you have the impression that at any time during the session the pen was being guide by an outside force?" Based on the results of Palmer (2011), it seems that the prediction should be that performance would be significantly better for Ps who experienced the OF from 1 to 40% of the time during the task than for other Ps (Hypothesis 1).

Based on data I had observed prior to looking at the AC scores, but before I observed any AC scores in the present study, I decided to modify the hypothesis to state simply that performance would be significantly better for Ps who have the experience that the associated hand is being moved by an outside force (for any percentage of the movements) during the task than for other Ps. There were two reasons for the switch. First, based on the results of Palmer (2011), I decided for the present study to interview Ps at the end of the session about their experience during the session. I had expected that the reason for the curvilinear effect in Palmer (2011) is that those who experienced the outside force < 40% of the time were in a qualitatively different state of consciousness than those who experienced it > 40% of the time. I found no evidence of such a difference in the interviews in the present study.

Second, a smaller percentage of the Ps said they had experienced an outside force in the present study (44%) than in the Palmer (2011) study (65%). This finding affected my decision to change the hypothesis for the following reason. Many of the Ps in the Palmer (2011) study were college students, whereas almost none were college students in the present study. Even though I was not the students' professor, I had a similar enough role as an "authority figure" that some of them might have wanted to please me, a common malady in mainstream psychology research called "demand characteristics." A good way to accomplish this objective would be to assure me that they had the "expected" and "hoped for" experience of their hand being moved by an outside force, even when it was not. To do this fully, they would not only want to acknowledge having the experience but to have it a high percentage of the time. I think more likely than an outright lie is adoption of an overly liberal conception of what the experience of an outside force would be like. This hypothesis can explain why 21% more Ps claimed the OF effect in Palmer (2011) than in the present study. If those Ps in the > 40% OF condition who presumably acquiesced to demand characteristics were shifted from that condition to the 0% force condition, better reflecting their true experience, the relationship would conform to the revised hypothesis for the present study.

Location *z*-scores. In the EQR condition, the 19 Ps who answered the OF question "yes" scored significantly above chance ($M = 0.78$, $SD = 1.22$), $t(18) = 2.79$, $p = .006$, one-tailed, and significantly higher than the 30 Ps who answered it "no" ($M = 0.19$; $SD = 1.01$); for the difference, $t(47) = 1.83$, $p = .037$, one-tailed. Thus, Hypothesis 1 was confirmed, but only for the EQR condition. On the other hand, there was a negligible, nonsignificant difference in the QL condition, where the six Ps who answered the OF question "yes" had an AC mean of -0.64 ($SD = 1.22$) compared to a mean of -0.82 ($SD = 0.84$) for the 13 who answered it "no," $t(17) = 0.37$, $p = .72$.

Number *z*-scores. How Ps answered the force question had no effect on the number hits. The 28 Ps who answered the OF question "yes" had a mean *z* of -0.06 (*SD* = 0.84) compared to a mean of -0.03 (*SD* = 1.10) for the 43 "no" responders. The difference is nonsignificant, $t(69) = 0.10$.

Trait Dissociation and AC

The Dissociative Processes Scale (DPS) used to measure trait dissociation has three subscales: Obliviousness, Imagination, and Detachment. Descriptive data for the three subscales are as follows. Obliviousness (*M* = 43.62, *SD* = 11.10); Imagination (*M* = 23.02, *SD* = 6.90); Detachment (*M* = 15.35, *SD* = 5.44). In our sample, the intercorrelations among the scales ranged from .44 to .63. The intention from the outset was to analyze the three scales individually. All the scales had acceptable skewness and kurtosis. The correlations of these scores with the AC measures are summarized in Table 2.

Location *z*-scores. *Obliviousness*. Hypothesis 2 predicted a significant positive correlation between location hits and scores on DPS Obliviousness in the EQR condition. This hypothesis was not confirmed, $r(47) = -.07$.

***Imagination*.** There was a significant positive correlation between location hits and DPS Imagination in the EQR condition, $r(47) = .35$, $p = .01$. The correlation was in the same direction and of comparable magnitude, although not significant, in the QL condition, $r(17) = .30$, $p = .21$.

***Detachment*.** There was a significant positive correlation between location hits and DPS Detachment in the EQR condition, $r(47) = .30$, $p = .03$. The correlation was in the same direction and of comparable magnitude, although not significant, in the QL condition, $r(17) = .25$, $p = .30$.

Number *z*-scores. *Obliviousness*. The correlation between DPS Obliviousness and number hits is nonsignificant in the EQR condition, $r(47) = -.03$, but surprisingly significant in the QL condition, $r(17) = .46$, $p = .05$.

Imagination. There was a significant positive correlation between number hits and DPS Imagination, $r(69) = .26$, $p = .03$. The correlation is nonsignificant in the EQR condition, $r(47) = .19$, but significant in the QL condition, $r(17) = .56$, $p = .01$.

Detachment. The correlation between number hits and DPS Detachment is nonsignificant in the EQR condition, $r(47) = .23$, $p = .11$, but significant in the QL condition, $r(17) = .62$, $p = .005$.

Trait Dissociation and OF Ratings

There is no significant difference between Ps who did and did not claim they felt their hand being moved by an outside force during the AC task on DPS Obliviousness. Thus, Hypothesis 4 was not supported. However, there is a substantial and highly significant difference in their scores on the Detachment subscale. The numbers are presented in Table 2.

Table 2
Means (Standard Deviations) Comparing Yes and No Responders to the OF Question on the DPS Subscales for the Entire Sample

Subscale	Yes (*N* = 35)	No (*N* = 45)	*t*	*p*
Obliviousness	45.26 (10.96)	42.71 (11.21)	1.02	.31
Imagination	23.34 (7.57)	22.78 (6.41)	0.36	.72
Detachment	17.86 (5.38)	13.40 (4.68)	3.96	.0002

Independence of Predictors

I wanted to know if each of the three significant predictors of location hits in the EQR condition

(OF question, DPS Detachment, DPS Imagination) could be demonstrated to have an effect independent of the other two. It is obvious from Table 2 that OF and Detachment are not independent, and that was confirmed by a regression analysis I performed with all three predictors. However, I wouldn't expect them to be independent, because they are respectively state and trait measures of the kind of dissociation that the OF responses were intended to reflect. Therefore, I created a new variable, *Dissociation*, that combines the two measures as if they were two items on a single scale. I coded a "yes" response to the OF question as +1 and a "no" response as -1, the dummy codes I used in the previous regression analysis. I then generated standard scores for OF and Detachment, added them, and divided by 2. This is the same way I created the location *z*-scores. For consistency, I also converted Imagination scores to standard scores. I then performed a new linear regression analysis with Dissociation and Imagination as the sole predictors. R for the model was .42, $F(2,46) = 4.97$, $p = .011$. The two betas were suggestively significant: Dissociation: $\beta = .24$, $p = .093$; Imagination: $\beta = .28$, $p = .052$. The same analysis gave a significant correlation between Dissociation and Imagination, $r(47) = .30$, $p = .034$. Finally, Dissociation is significantly correlated with location hits, $r(47) = .35$, $p = .013$.

Discussion

The initial analysis revealed significant psi-missing in the QL condition (AC responses made with the left hand while reading quotations on the computer screen) offset by significant psi-hitting in the other three conditions, which were pooled to create the EQR condition. The significant psi-missing in the QL condition makes sense in that reading quotations is more cognitively demanding than blanking the mind, and any psychomotor task is more difficult to perform with the nondominant hand. Combining the two added yet another layer of difficulty. Although I attempted to mitigate this problem by giving Ps liberal practice under the conditions they would experience in the formal task, as I explained earlier, I never expected that it would be enough to completely solve the problem. These difficulties can easily lead to frustration, and such a negative mindset has long been associated with psi-missing in parapsychology.

The OF question was intended as a measure of a dissociated state during the AC task. Its validity in this regard was assessed by determining whether it correlates significantly with an appropriate trait measure of dissociation capacity, for which I chose the DPS. Specifically, I expected the OF question to correlate with the Obliviousness subscale of the DPS, and this expectation was reflected in Hypotheses 2 and 4.

However, there was not a significant relationship between the OF question and DPS Obliviousness. Instead, there was a quite strong positive relationship between the OF question and another of the DPS subscales, Detachment. Although not predicted, this relationship makes sense. Watson (2003, p. 300) defines the Detachment subscale as measuring "feelings of depersonalization and derealization," which doesn't seem very relevant. However, greater relevance can be found if one examines the six items that compose the subscale. At least three of these reflect separation from the body, very much like an out-of-body experience. For example, Item 17 reads "At times I have felt disconnected from my body." It is not a huge inferential leap to extend this concept to one's hand being outside of, and not controlled by, the rest of one's body and the mind that governs it. It is striking that the correlations between the OF question and the other two DPS scales, Imagination and Obliviousness, are nowhere near significant, despite the fact that the three subscales correlated with one another in the .4 to .6 range. Finally, DPS Detachment scores were directly and positively related to location hits in the EQR condition. In short, the present study confirms that trait dissociation is indeed reflected by the tendency to experience one's hand being moved by an OF during an AC motor task.

The reinterpreted-as-positive relationship I found in my previous motor automatism study (Palmer, 2011) between AC hits and perception of the hand as being moved by an outside force at least some of the time during the AC task was found in the present study, confirming Hypothesis 1, albeit only for the EQR condition. Surely there is no reason for the effect to hold in the QL condition, where there was significant overall psi-missing. If anything, one would predict a reversal of the OF/AC relationship in this condition, for a reason I discuss further below. One could say that the significant correlation between location hits and the Dissociation measure created for the regression analysis provides a conceptual confirmation of Hypoth-

esis 1, because the combination of OF and DPS Detachment (the components of Dissociation) is a better measure of the dissociation construct that the OF question was intended to capture than the OF question alone. Also, this correlation is significant by a two-tailed test.

The significant correlation of AC hits with DPS Detachment in the EQR condition was matched by a significant correlation of AC hits with DPS Imagination. Also, both DPS scales correlated positively and significantly with *number* hits in the entire sample. For number *z*-scores there was not the sharp difference between total scores in the EQR and QL conditions found with location *z*-scores. For both predictors, the correlation with number *z*-scores was actually stronger in the QL condition than in the EQR condition.

An Overall Interpretation: The Two-Process Model

So, what might this complex pattern of results be telling us? First, note that accessing a number as such requires cognitive activity in the strong sense. Whereas such cognitive processes could also be used to access square and quadrant, they are not necessary. If Ps were able to follow the instructions properly, any such cognitive activity should have been unconscious. Second, recall that Watson (2003, p. 300) describes the Imagination subscale as assessing "absorption, imaginativeness, and fantasizing." This describes activity that is more cognitive than for the other two scales, even to the point of reflecting imagery.

My interpretation from all this is that there were two kinds of psi processes contributing to the AC test outcomes, and I will label what follows as my two-process model. One process was almost purely motor (which can be expressed metaphorically as the hand performing the AC task) and the other was primarily cognitive. The low-magnitude results parapsychologists consistently find in AC tests suggests that most individual responses to targets are not psi-mediated; they are simply "wild guesses." This leaves plenty of room for different processes to mediate the responses on different trials in a forced-choice AC test such as the one used in this experiment. In the present case, that would mean that within a 36-trial run exhibiting some psi, most trials would be chance, a few would be motor-psi-mediated and a few others cognition-psi-mediated. It also assumes, as noted above, that the variance associated with DPS Detachment can be divided into two components: that which significantly predicted responses to the OF question (Component A) and that which correlated significantly ($r = +.52$) with DPS Imagination (Component B).

The motor process was active only in the EQR condition and affected only location hits. It was predicted by Component A of DPS Detachment and was experienced by Ps who felt their hand being moved by an OF. This model is supported by the following specific findings: (a) the significant positive relation between DPS Detachment and the OF question, along with the lack of a significant relation between the OF question and DPS Imagination; (b) the significant correlation between the OF question and location *z*-scores exclusively in the EQR condition; (c) the significant correlation of DPS Detachment with location *z*-scores only in the EQR condition; (d) the lack of a significant relation between number *z*-scores and either DPS Detachment or the OF question in the EQR condition.

The cognition process was operative in all the conditions, including QL. It was predicted primarily by DPS Imagination, and only by DPS Detachment insofar as that was correlated with DPS Imagination (Component B). The model is supported by the following specific findings: (a) a significant correlation between DPS Imagination and location *z*-scores in the EQR condition; (b) a significant correlation between DPS Imagination and number *z*-scores in the entire sample. Although either the motor or the cognitive process could be used for location hits, only the cognitive process can achieve number hits; the ability of the "hand" to point to the correct spot on the grid is useless for identifying numbers per se.

The two-process model is also supported by the regression analysis. In fact, seeing if such support could be found was my main motive for performing this analysis. I have a philosophy about the respective roles of bivariate and multivariate relationships that is the reverse of what seems to be the norm in psychology and to some degree in parapsychology: In my view, bivariate relationships should be primary, and targeted (as opposed to global) multivariate analyses should be used to interpret their meaning. I think what follows provides a good example of why this approach makes sense. First, the suggestively significant betas for Dissociation and Imagination provide support for the proposition that these variables made independent

contributions to the prediction of location hits in the EQR condition. On the other hand, the significant correlation between Dissociation and Imagination, combined with the significance of the full model, provides support for the complementary proposition that these variables also made joint contributions to the prediction. The suggestive confirmation of the independent contributions of Dissociation and Imagination is more impressive than it might seem on the surface because it was found despite the .3 correlation between them. Given the notoriously poor reliability of forced-choice ESP scores, a suggestive result is the best that one could reasonably expect if the two-process model is correct.[2]

Some additional comments are needed about number hits in the QL condition. The correlation between number *z*-scores and DPS Imagination was particularly strong in the QL condition. However, what is most surprising is that the direction of the relationship was positive, because, as I noted earlier in the paper, one would expect a negative correlation when the AC mean is below chance (Palmer, 1975). My explanation of the positive correlations between the DPS scales (even Obliviousness) and number *z*-scores in the (psi-missing) QL condition draws on my theory of different processes operating on different trials. Specifically, high-Imagination Ps were able to overcome on some trials the apparent frustration manifested on other trials by the difficulty of the QL procedure, but only for cognitive responding. In other words, the high-Imagination Ps were able to partly offset or cancel the psi-missing trials caused by frustration with psi-hitting trials mediated by their relatively good cognitive psi abilities, bringing the mean number *z*-score obtained by these high-Imagination Ps close to chance. This offset would not apply to low-Imagination Ps, so their scores on number hits would remain low, the net result being a positive correlation between DPS Imagination and number hits in the QL condition. Why Ps could not do that using the motor process remains unclear.

Finally, the two-process model has caused me to recognize that a cognitive process was probably involved in producing the results of the Palmer (2011) study, something I had previously been in denial about. Specifically, the results in Palmer (2011) were best when participants based their ESP guesses on a combination of their alphabet board responses and mental imagery they might recall from the session.

Investigator Psi

Any psi researcher is required in my view to at least acknowledge the possibility that their results were contaminated to at least some degree by investigator psi, or I-psi. (I use this term rather than the more customary E-psi because I think the potential psi sources should include the Principal Investigator, who may not be interacting with the participants.) Although I-psi can never be ruled out, it can be minimized by eliminating as much as possible intuitive or random investigator decisions. This is why I had Kruth use the seed numbers 1 and 2 to activate the algorithm for target decision. It was my original intent to have the condition assignments made by a continuation of that number sequence, but due to logistical complications it was in fact done by my programmer Ruszaj using another algorithm. He had no stake in the outcome of the study and showed no particular interest in its conceptual aspects, so I doubt he had the motivation I consider necessary to be an implicit psi source. In theory I could have influenced the process myself, but at the time I had no idea when Ruszaj would perform the task or even that he was going to be the one to do it.

By far the most likely source of I-psi in the study was my adjustment of the response sequences for some Ps in the eyes-closed conditions. Even though I did my best to make these decisions rationally, more than one option could occasionally be defended on rational grounds, and it is possible that psi could have tilted my decision process in a favorable direction in some cases. Clearly, the results in these conditions were better for Model 2 than for Model 1. However, there were equally strong results in the QL condition, where the impact of my decision-making was minimal. On the other hand, my general expectation in studies such as this is that both the participants and the investigator have input into the results (Palmer & Millar, 2015), and I see no reason to make an exception in this case.

[2] The more general proposition that underlies my model for this experiment, namely that different psychological processes mediate psi in different trials of the same run, is to my knowledge novel in parapsychology. Thus, I think it is worth presenting even if it cannot be considered to have been conclusively confirmed.

Conclusion

First, to meet my own requirement for publication of empirical reports in the *JP*, I hereby declare that all significant results from this exploratory study must be considered tentative unless or until they are replicated, unless they are replications themselves. The only relationship that meets the latter standard is that confirming Hypothesis 1, but only if one accepts my modification the hypothesis.[3]

On the other hand, although only one of the four hypotheses was confirmed by the data (state-dissociation-AC), and in this case the hypothesis had to be reworded (for reasons I consider justified), there was a surprisingly large number of significant effects that were not hypothesized, most of which make sense both individually and in relation to one another, and they could have been justifiably hypothesized were I more astute. I maintain that more attention should be paid to whether an effect is *hypothesizeable* than whether it is *hypothesized*.[4] By this standard, and granted that little if anything was *established* by this study, it seems to me that the likelihood is sufficiently great that the psychological processes functioning as described in the two-process model contributed to the production of weak but genuine psi effects in this study that the model could profitably be taken into account in designing future process-oriented research of relevance to the model. For my part, I have completed a followup study in which I attempted to train the best individual Ps in the present study to develop proficiency in the procedure through extensive practice with subliminal auditory feedback of hits. The purpose was to increase the reliability of their AC scoring. The results will be reported in a separate paper.

References

Bartlett, J. (2002). *Bartlett's familiar quotations: A collection of passages, phrases, and proverbs traced to their sources in ancient and modern literature* (17th ed.). Boston, MA: Little, Brown.

Braud, W. G. (1975). Psi-conducive states. *Journal of Communication*, *25*, 142–152.

Hilgard, E. (1977). *Divided consciousness*. New York, NY: Wiley.

Maher, M., & Schmeidler, G. R. (1977). Cerebral lateralization effects in ESP processing. *Journal of the American Society for Psychical Research*, *71*, 261–271.

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

Metcalf, F. (2009). *Dictionary of modern humorous quotations* (3rd ed.). New York, NY: Penguin.

Millar, B. (1978). The observational theories: A primer. *European Journal of Parapsychology*, *2*, 304–332.

Myers, F. W. H. (1903). *Human personality and its survival of bodily death* (2 vols.). London, UK: Longmans, Green.

Palmer, J. (1975). Three models of psi test performance. *Journal of the American Society for Psychical Research*, *69*, 333–339.

Palmer, J. (2001). A mail survey of Ouija board users in North America. *International Journal of Parapsychology*, *12*, 67–93.

Palmer, J. (2011). Motor automatisms as a vehicle of ESP expression. *Journal of Parapsychology*, *75*, 45–60.

Palmer, J. (2013). *JP* publication policy: Statistical issues [Editorial]. *Journal of Parapsychology, 77*, 5–8.

Palmer, J. (2016). Statistical issues in parapsychology: Hypothesis testing—plus an addendum on Bierman et al. (2016) [Editorial]. *Journal of Parapsychology*, *80*, 141–143..

Palmer, J., Bogart, D. N., Jones, S. M., & Tart, C. T. (1977). Scoring patterns in an ESP ganzfeld experiment. *Journal of the American Society for Psychical Research*, *71*, 122–145.

Palmer, J., & Millar B. (2015). Experimenter effects in parapsychological research. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st Century* (pp. 293–300). Jefferson, NC: McFarland.

Persinger, M. A., & Makarec, K. (1993). Complex partial epileptic signs as a continuum from normals to epileptics: Normative data and clinical populations. *Journal of Clinical Psychology*, *49*, 33–45.

Schmidt, H., & Pantas, L. (1972). Psi tests with internally different machines. *Journal of Parapsychology*, *36*, 222–232.

[3] This standard is my (more demanding) substitute for multiple analysis corrections, to which I have strong objections (Palmer, 2013).

[4] For a further defense of this approach, see Palmer (2016).

Solomons, L. M., & Stein, G. (1896). Studies from the Psychological Laboratory of Harvard University. II. Normal motor automatism. *Psychological Review*, *3*, 492–512.
Tellegen, A. (1978). *Differential Personality Questionnaire*. Minneapolis, MN: University of Minnesota.
Watson, D. (2003). Investigating the construct validity of the dissociative taxon: Stability analyses of normal and pathological dissociation. *Journal of Abnormal Psychology*, *112*, 298–305.

Rhine Research Center
2741 Campus Walk Ave., Bldg. 500
Durham, NC, USA
john@rhine.org

Acknowledgement

I am grateful to the Bial Foundation for their support of this research.

Abstracts in Other Languages

French

COGNITION ANOMALE, DISSOCIATION, ET AUTOMATISMES MOTEURS

RESUME : 80 participants ont effectué une tâche de cognition anomale reflétant un automatisme moteur, précédée par un exercice de relaxation progressive. Une grille de 4x4 de 16 carrés d'un pouce de côté contenant chacun un nombre entre 1 et 4 était colléée sur une tablette informatique. Lors des 36 essais, les participants ont exploré la grille avec un stylet, enregistrant leurs réponses dans les carrés cibles sélectionnés aléatoirement en se fixant pendant 1 s. Ils étaient assignés dans quatre groupes dans un protocole à 2x2 conditions avec les variables indépendantes de la latéralité manuelle (droitier ou gaucher) et du facilitateur de dissociation (yeux fermés et vide de l'esprit contre lecture de citations sur un écran). La variable dépendante était le lieu des succès, une combinaison non pondérée de carré standardisé et de succès par quadrants. Une analyse ANOVA a révélé un psi-missing significatif dans la condition [citations/gaucher] et un psi-hitting signficatif dans les autres conditions (EQR). Des scores élevés dans la composante Détachement (DET) de l'échelle des processus dissociatifs (DPS) et les témoignages de main semblant se mouvoir grâce à une force extérieure (OF) durant la tâche de cognition anomale prédisaient conjointement les succès de localisation dans la condition EQR, comme le faisait également le score d'Imagination (IMA) issu de la DPS. IMA et DET étaient corrélés significativement avec les succès dans toutes les conditions. Cette étude exploratoire fut interprétée comme reflétant la médiation du psi par un processus moteur et un processus cognitif au cours de différents essais.

German

ANOMALE KOGNITION, DISSOZIATION UND MOTORISCHE AUTOMATSMEN

ZUSAMMENFASSUNG: 80 Freiwillige absolvierten eine anomale Kognitionsaufgabe in Form eines motorischen Automatismus, der eine progressive Entspannungsübung voranging. Auf einem Tabletschreibcomputer war ein 4x4-Gitter von 16 einzolligen Quadraten aufgetragen, von denen jedes die Zahlen 1-4 enthielt. Bei 36 Versuchsdurchgängen erkundeten die Probanden (Pbn) das Gitter mit einem Computerstift, wobei ihre Reaktionen auf den zufällig ausgewählten Zielquadraten aufgezeichnet wurden, sobald sie auf der 1 stoppten. Sie wurden 4 Zellen in einem 2x2-Design zugeordnet, wobei die unabhängigen Variablen die verwendete Hand (links oder rechts) und ein Dissoziation-Unterstützer (Gedankenleere mit geschlossenen Augen vs. Lesen von Zitaten auf einem Schirm) waren. Die abhängige Variable war die Lokalisierung der Treffer, eine ungewichtete Kombination des standardisierten Quadrats und der Quadrantentreffer. Eine

ANOVA ergab ein signifikantes Psi-Missing in der Zitate/links-Bedingung und ein signifikantes Psi-Hitting an anderer Stelle (EQR). Hohe Scores auf der Detachment (DET)-Komponente der Dissociate Process Scale (DPS) und Berichte, die Hand sei durch eine Äußere Kraft (ÄK) während der AK-Aufgabe bewegt worden, sagten zusammengenommen hohe Trefferscores bei der Lokalisierung in der EQR-Bedingung voraus, wie es auch bei der DPS-Imagination (IMA) der Fall war. IMA und DET korrelierten signifikant mit Treffern auf der Zahl über alle Bedingungen hinweg. Dieser exploratorischen Studie liegt die Interpretation zugrunde, dass Psi durch einen motorischen Prozess und einen kognitiven Prozess bei verschiedenen Versuchen vermittelt werden kann.

Spanish

COGNICIÓN ANÓMALA, DISOCIACIÓN, Y AUTOMATISMOS MOTRICES

RESUMEN: Ochenta voluntarios completaron una tarea de cognición anómala que refleja automatismo motriz, precedida por un ejercicio de relajación progresiva. Se usó, adherida a una tableta de escritura de computadora, una cuadrícula 4x4 de 16 cuadrados de 1 pulgada cada uno con números del 1-4. Durante 36 pruebas, los participantes exploraron la cuadrícula con la pluma de la computadora, registrando sus respuestas parando durante 1 s sobre los cuadrados que pensaban eran los objetivos y que habían sido seleccionado al azar. Estas variables independientes fueron asignadas a 4 celdillas en un diseño 2x2: mano usada (derecha o izquierda) y un facilitador de disociación (poner la mente en blanco con los ojos cerrados o leer citas en la pantalla). La variable dependiente fue la localización de los aciertos, una combinación estandarizada neta de aciertos en las celdillas y cuadrantes. ANOVAs revelaron una falla significativa en la combinación citas/mano izquierda y aciertos significativos en el resto de las condiciones (EQR). Las puntuaciones altas en el componente de Desapego (DET) de la Escala de Procesos Disociativos (DPS) y experiencias de la mano movida por una fuerza externa durante la tarea predijeron conjuntamente puntuaciones altas de localización en la condición EQR, así como el componente de Imaginación de la DPS (IMA). IMA y DET correlacionaron significativamente con los aciertos en el número en todas las condiciones. Este estudio exploratorio puede interpretarse como psi mediada por un proceso motriz y un proceso cognitivo en diferentes pruebas.

Journal of Parapsychology, 81(1), 63-72

EXPERIMENTER FRAUD: WHAT ARE APPROPRIATE METHODOLOGICAL STANDARDS?

By J. E. Kennedy

ABSTRACT: Discussions of experimenter fraud in parapsychology have missed a key lesson from the Levy case. The standard procedure for handling scientific fraud is an after-the-fact (post hoc) investigation. Post hoc investigations cannot be expected to be effective in parapsychology because signs of fraud in the data can be attributed to psi, as happened with Levy. In parapsychology, compelling evidence of fraud usually requires direct covert detection of fraud as it occurs during an ongoing experiment, as in the Levy case. However, such covert measures by colleagues are not a practical strategy for addressing fraud and are not expected in other areas of science. The standard that experimental procedures should make fraud by one experimenter very difficult or impossible has long been advocated in parapsychology but has not been implemented in recent decades. This standard was implemented in my experience working in regulated medical research and should eliminate the vast majority of cases of fraud—which start when one experimenter finds data manipulation or fabrication easy and tempting with very little possibility of detection. This standard provides a systematic and effective way to address experimenter fraud and should become part of the new standards for research in the behavioral sciences.

Keywords: experimenter fraud, experimenter misconduct, research methodology, research standards, confirmatory research

The recent extensive discussions of methodological issues for psychological research have included the topic of experimenter fraud (John, Loewenstein, & Prelec, 2012; Simonsohn, 2013; Strobe, Postmes, & Spears, 2012). In extending these discussions to parapsychology, Stokes (2015) and Kennedy (2014) raised the possibility that experimenter fraud may be more extensive than is generally realized and may significantly compromise the research findings, particularly when combined with other forms of methodological bias. Palmer (2016) and Roe (2016) challenged these points.

Virtually all discussions of fraud in parapsychology describe the Levy case as the most definitive evidence of experimenter fraud. As one who was involved in the Levy exposé, my observation is that the various discussions miss the key factors that were involved and the important implications of the Levy case. In addition, writers often make surmises about what happened without direct knowledge. For example, Rogo's (1985) description contains numerous errors and is not a reliable source of information (Kennedy, 2017).

In this paper I discuss key points and implications of the Levy case that have not been discussed before and describe my experiences working in research environments where measures to prevent fraud were standard procedure. I also make recommendations for dealing with fraud based on my experiences combined with the findings from published studies of fraud.

As a frame of reference, it is useful to describe three alternative positions or expectations with regard to experimenter fraud.

1. The expectation of *fraud-proof experiments* is based on the assumption that research can and should be conducted in a way that completely precludes the possibility of experimenter fraud. As Palmer (2016) pointed out, this is the standard for extreme skeptics such as Hansel (1966, 1980). Palmer also pointed out that this expectation is currently widely rejected in science. For me this expectation is eliminated by the fact that no measures could prevent collusion among experimenters to produce fraudulent results.

2. The expectation that *fraud by one experimenter should be very difficult or impossible* is based on the recognition that the vast majority of cases of fraud start when an experimenter is alone with the data and finds that data manipulation or fabrication is easy and tempting with very little possibility of detection. This standard has long been recommended in parapsychology (Akers, 1984; Dalton et al., 1996; Rhine, 1974, 1975) and was a methodological standard in my experience working in regulated medical research (described below).

3. The expectation that *procedures to prevent experimenter fraud are unnecessary* is usually based on the argument that independent replication will reveal and overcome fraud. Palmer (2016) raised a related argument that procedures to prevent experimenter fraud could create a paranoid work environment that should be avoided. He also argued that discussion or modeling of possible fraud for specific studies are implicit accusations of fraud and therefore are not ethical in the absence of strong evidence that fraud has occurred.

Background on Experimenter Fraud

Experimenter fraud is an established factor in scientific research (Broad & Wade, 1982; John et al., 2012; Stroebe et al., 2012; Retraction Watch, n.d.). The extent of occurrence of fraud is unknown because undetected instances are likely and institutions have often been reluctant to make cases of fraud known publicly.

Independent replication and peer review for publication have generally not been effective at detecting even extensive fraud (Broad & Wade, 1982; Strobe et al., 2012) and do not pose a significant risk of detection for those contemplating fraud. The primary symptom of fraud is inconsistent results among experimenters, but such differences are virtually never attributed to fraud.

A recent analysis of cases of scientific fraud reported that most frauds are detected by whistleblowers inside an organization and that "fraudsters are usually reluctant to make available the data they allegedly collected" (Strobe et al., 2012, p. 682). The authors noted that "whistleblowers are likely to remain the single most effective instrument against scientific cheating" (p. 682). This recent analysis confirmed the same basic points made 30 years earlier by Broad and Wade (1982).

The normal process for handling experimenter fraud is an investigation by a committee after suspicions of fraud have been formally raised (Gross, 2016; Strobe et al., 2012). The committee examines publications and asks for raw data and other research records. Evidence or signs of fraud are typically found in the data and publications, including inconsistencies, data patterns that are artifacts of fraud, and/or data that are "too good to be true" (Strobe et al., 2012). Investigations of fraud are expected to take 10 months if all goes smoothly, but in practice, longer times have been common (Gross, 2016).

Gross (2016, p. 700) observed that "there appear to be no systematic empirical studies of the characteristics of perpetrators of scientific misconduct and no good evidence for any common characteristics." He pointed out that the cases that get extensive publicity usually involve highly ambitious researchers who rise rapidly in elite institutions. However, these highly publicized cases cannot be assumed to be representative of all cases of experimenter fraud.

Three categories of fraud can be distinguished: detected, suspected, and undetected. In cases of detected fraud, initially suspected or observed fraud is investigated and unambiguously resolved as fraud. In cases of suspected fraud, the evidence of fraud is not fully resolved, even though apparent fraud may have been observed by a colleague. Suspected fraud includes cases that are not investigated and remain at the level of rumor as well as cases that are reported and investigated but have inadequate evidence to determine whether fraud did or did not occur. Undetected frauds are cases that do not reach the point of suspicion by colleagues. Reliable data obviously cannot be obtained about undetected fraud.

Surveys have been conducted asking scientists about admitted, observed, or suspected fraud (Fanelli, 2009; John et al., 2012). The accuracy of the findings is questionable for such surveys because the respondents may be biased about this topic. The generalizability of the samples is also questionable. In addition, the surveys cannot address undetected fraud.

However, the surveys may provide insights about the rate at which cases of suspected fraud are reported, investigated, and resolved. In commenting on one of the more methodologically sound surveys, Titus, Wells, and Rhoades (2008) stated "Extrapolating the survey results projects an alarming picture of under-reporting" (p. 981) They argued that all research centers should have the policy that any suspected researcher misconduct must be reported and must be thoroughly and fairly investigated

.

My Experiences Exposing Experimenter Fraud

As noted above, I was involved in exposing the fraud of W. J. Levy. The experiment in which Levy was exposed was officially my experiment, and Levy was my co-experimenter. I also had the leading role in investigating the extent of his fraud for three lines of research (Kennedy, 1975a, 1975b, 1975c). Contrary to the incorrect comments by Palmer (2016), Doug Stokes did not have a role in the Levy exposé.

The Levy exposé was different from the typical case of exposed scientific fraud because direct evidence of fraud was obtained as the fraud occurred. Jim Davis, Jerry Levin, and I established a hidden recording of the output of the RNG used in the experiment before the point in the circuit where Levy pulled a plug to introduce bias. Recordings were made during an actual experiment without Levy's knowledge. Davis also covertly observed the equipment during the period Levy was pulling the plug.

As noted above, suspicions of scientific fraud are normally handled by an investigation after the fact (post hoc) without direct experimental evidence as the fraud occurs. For the 40 cases of fraud summarized by Strobe et al. (2012), only one is described as a "sting operation" in which a colleague trapped a fraudulent researcher, as we did in the Levy case.

Evidence of fraud from post hoc investigations will usually be unconvincing in parapsychology because a fraudulent researcher can claim that the signs of fraud in the data are actually psi effects. The most conspicuous artifacts of Levy's fraud were long strings of consecutive hits that had extremely low probability of occurring by chance (Kennedy, 1975b). Levy presented these strings as psi effects, and this interpretation had become accepted at the lab. Within the worldview of parapsychology, the claim that signs of fraud are actually psi effects is nearly irrefutable and cannot be resolved by post hoc analysis. For other areas of science, fraudulent researchers do not have psi as a virtually indisputable alternative explanation for signs of fraud in the data.

For the Levy case, we needed to obtain direct, experimental evidence of fraud as it was occurring to establish that the effects were not due to psi. We expected that accusations without such evidence would lead to a subsequent post hoc investigation that would produce a prolonged, intense debate with an inconclusive outcome. Evidence of fraud found in post hoc analysis would be considered to be possible psi effects. In the end, the negative impacts for the accusers would be as great as or greater than for the accused. Before we openly raised the issue of fraud, we needed to have indisputable evidence that some fraud had actually occurred.

Covertly obtaining direct definitive evidence as fraud is being conducted will usually be necessary for resolution of fraud in parapsychology but is generally not a practical goal. The effort to obtain such evidence is beyond what is reasonable in a professional setting. The need to maintain normal interactions with a close colleague while covertly planning and conducting steps for his exposure and resulting ruined career requires a degree of acting and compartmentalization that many scientists do not have. For me it was very difficult. Many pivotal decisions had to be made quickly in secrecy and under stress. In addition to deciding the strategy and technical details for collecting unequivocal evidence, multiple people needed to be involved to establish overwhelming credibility. If it was even remotely feasible, Levy could claim the accusations were false and based on fabricated data. Decisions had to be made about who could handle the acting and extreme secrecy, how they should be approached, the risks of possible compromising communication, and the roles for the various people. In addition, it was sometimes necessary to deceive colleagues in order to keep the preparations secret.

These distasteful steps were necessary to resolve the matter unambiguously rather than creating an irresolvable situation with suspicions but no compelling evidence, as has occurred for other cases of

suspected experimenter fraud in parapsychology. In fact, such a situation had occurred previously with Levy. When Jerry Levin first observed Levy behaving suspiciously near some wires that could be used to manipulate the results, Jerry responded by covering the wires with tape to prevent potential fraud. Jerry did not clearly observe fraud and had only suspicions. However, taping the wires let Levy know that he was suspicious and effectively eliminated the possibility of resolving Levy's fraud in the line of research Jerry was conducting.

I found Jerry's suspicions to be unconvincing and dismissed them—until I later observed Levy apparently manipulating data in another line of research. My observations would have been adequate to initiate an investigation but did not provide the type of indisputable evidence that would be needed to overcome Levy's counterclaims that the accusations were mistaken or fabricated and that the effects were actually due to psi. Carefully planned, indisputable experimental evidence was needed as Levy actually manipulated the data, and multiple people needed to be involved.

Based on my experience exposing fraud, I think it is very unlikely that instances of experimenter fraud in parapsychology will be convincingly resolved. Obtaining convincing evidence of fraud in parapsychology is much more difficult than in other areas of science because the normal process of conducting a post hoc investigation will usually not be effective. Signs of fraud can easily be explained away as psi effects in parapsychology, but not in other areas of science. Another case in which data analyses found patterns that would normally be construed as signs of data manipulation but are ambiguous if PK is considered plausible is described in Kennedy (1980a, 1980b). Compelling evidence during the actual manipulation of the data—a sting operation—is needed to establish that the effects were not due to psi. However, that typically requires covert effort that is not practical for scientific research.

It is usually much easier to avoid dealing with experimenter fraud than to make the effort to fully resolve the matter. Even when clear evidence of fraud is found, the effort to deal with the fraud is very time-consuming and distracts researchers from their main interests. The investigation of the extent of Levy's fraud took about a year, which, as noted above, is common for investigations of fraud. In addition, the adverse effects for the work environment are often significant. For the Levy case, the exposé would clearly create major disruption of the work environmental at the beginning of the summer study program that Levy had organized. This disruption would be very detrimental for everyone at the lab, including those exposing Levy. In fact, the initial reaction of one of the three people involved in the exposé was to suggest that a long-term discrete investigation be conducted for several months or longer that would not disrupt the work environment, particularly over the summer. The other person and I vetoed that idea.

Broad and Wade (1982) argued that it is likely that only the most extreme, careless frauds have been detected. That conclusion is consistent with the experience in parapsychology. Levy's fraud appears to have become pervasive and irrational. For one experiment, Levy published fabricated results even though the original data and analysis programs were stored on backup tapes and provided completely different results (Kennedy, 1975a). Those of us involved in the exposé and subsequent investigations did not anticipate such irrational behavior.

For the four major lines of research Levy had conducted, fraud was exposed as it occurred in one and the data for a published study was clearly fabricated in another (Kennedy, 1975a). Strong circumstantial (post hoc) evidence of fraud was found in the other two lines of research (Kennedy, 1975b). Those skeptical of psi will interpret the circumstantial evidence as unequivocal evidence that fraud occurred in those studies. However, those of us involved in the Levy exposé believed that circumstantial evidence alone would not be accepted within parapsychology as compelling evidence that the effects were due to fraud rather than due to psi as claimed by Levy.

My Experiences Preventing Experimenter Fraud

My attitudes toward experimenter fraud have also been influenced by about 20 years of work in medical research. In my experience in regulated medical research, measures to prevent unintentional or intentional (fraudulent) data alterations were an accepted part of the research culture. In pharmaceutical

research, regulatory agencies audit key sites where data are collected and processed. I managed the software infrastructure for data management and analyses at a company and was the first person the FDA auditor wanted to interview. The auditor asked about every significant step in the development, validation, and use of the software systems and repeatedly asked what steps were taken to verify that unintentional or intentional data alterations did not occur.

For example, after learning that a laboratory transferred certain data electronically and a programmer imported and reformatted the data, the auditor asked "How do you know the programmer did not change the data?" The questions were carefully phrased to include both intentional and unintentional data changes. I explained that the laboratory sent another copy of the data directly to another person, and a third person compared that copy to the electronic data output by the programmer. Of course, we had documentation for that comparison. The auditor did not ask about possible errors by the person checking the data or about collusion between the programmer and the person checking the data.

The auditor assumed that intentional or unintentional data alterations by one person should be difficult or impossible. Two independent copies of key data can meet this criterion and provide an important level of confidence when research findings are challenged.

Double-checking a colleague's work is standard procedure in pharmaceutical research. A surprising number of mistakes are discovered. Regulatory auditors expect documentation of this double-checking. These verifications are an established part of the research culture and are not interpreted as questioning a person's integrity or competence. When the costs of making a mistake are high, people want their work verified.

I found that working in an environment with routine practices to prevent fraud was much preferable to my experiences in parapsychology. In fact, the strategy for exposing Levy involved duplicate records and a colleague observing Levy's actions during the experiment. These are the same basic procedures that are used to prevent fraud. In research environments with open efforts to prevent fraud (and also prevent unintentional errors), these procedures are expected and are considered good methodology. However, in environments without such measures, undetected fraud can be easy and tempting, and discussion of these practices can be considered inappropriate implicit accusations of fraud or incompetence.

If I would have told the auditor that I considered questions about the integrity of the programmer to be unethical and inappropriate, and that I assumed the programmer did not change the data and believed that efforts to verify that assumption created a bad, paranoid work environment, I would have failed the audit and been fired—appropriately so. Those arguments were not viable in the research culture.

Confronting Experimenter Fraud

The first and most fundamental question is whether the research culture allows the topic of experimenter fraud to be discussed without being taken as personal accusations. More generally, "for the ideologists of science, fraud is taboo, a scandal whose significance must be ritually denied on every occasion" (Broad & Wade, 1982, p. 142). These types of idealistic arguments are no longer viable. "As unpalatable as it is, to complete the culture change initiated in the second half of last century, we have to accept the fact that fraud can happen in our midst and that we have to look out for it" (Stroebe et al., 2012, p. 684).

In a healthy research environment experimenter misconduct (fraud and biased methodological practices) is considered an appropriate and necessary topic of discussion, including about specific studies by specific experimenters. Such discussions reflect a high priority on good methodology and must not be taken as personal accusations.

The next question is: Can we accept that independent replication and peer review generally are not effective at detecting or deterring fraud? Stroebe et al. (2012) and Broad and Wade (1982) reached that conclusion in their studies of fraud, and it has been true for the two prominent cases of fraud in parapsychology (Levy and Soal—see Beloff, 1993, for a description of the Soal case). In fact, the long-established finding in parapsychology of consistent differences among experimenters could be taken as a symptom of experimenter

misconduct (fraud and/or biased methodology). The exclusion of such considerations when discussing these experimenter differences brings into focus how ineffective independent replication is for deterring or detecting fraud. As is clear from Stroebe et al. (2012), replication is at best an extremely inefficient, slow, and costly strategy for dealing with fraud and does not deter fraud. The resources required to conduct well-powered confirmatory studies are usually substantial. The resources for conducting multiple confirmatory studies of a fraudulent finding will often be a significant diversion of the limited resources available for behavioral research. A researcher may initially rationalize fraud as necessary to obtain funding for a more effective research program—which was one of Levy's explanations for his fraud. Unsuccessful independent replications will virtually never be identified as indicating fraud, and thus an experimenter contemplating fraud need not be concerned about the threat of detection.

One of the most important questions is: Can we accept the fact that we simply do not know how much undetected experimenter fraud actually occurs? Cases of detected fraud are a proportion of the total cases of actual fraud and the magnitude of undetected fraud is unknown. The common arguments that detected fraud is rare in parapsychology and occurs at the same or lower rates as in other areas of science (Bierman, Spottiswoode, & Bijl , 2016; Broughton, 1991; Roe, 2016) provide no useful conclusions about the occurrence of undetected fraud in parapsychology or in other areas of science. Bierman et al. (2016) ignored undetected fraud and considered only detected, suspected, and admitted fraud in their simulations—which therefore underestimated the actual occurrence of fraud by an amount that is unknown.

Stroebe et al. (2012, p. 682) commented that the cases of detected fraud in their report "are likely to be the tip of an iceberg of fraudsters." Titus et al. (2008) made a similar point. Broad and Wade (1982) acknowledged that the actual rate of experimenter fraud is unknown, but thought it is likely that for every case of major detected fraud, "a hundred or so go undetected" (p. 87). These speculations do not provide strong conclusions, but they do indicate the magnitude of the uncertainty.

Stokes (2015) and I think it is likely that a substantial amount of undetected fraud has occurred in parapsychology and in psychology given past research practices. Palmer (2016) and Roe (2016) argued that these are speculations without convincing evidence. However, they also did not provide convincing evidence that substantial undetected fraud has not occurred. Their papers focus on detected fraud and may give readers the impression that they believe the possibility of undetected fraud in past research can be ignored—which was my definite impression from their papers.

If research was conducted with measures to prevent and to detect experimenter fraud, the argument that undetected fraud is negligible would be plausible. However, in the absence of such measures, I do not see a scientific basis for this argument. In personal communication (September 12, 2016), Palmer emphasized that he did not intend to argue that undetected fraud can be ignored. He believes there is not convincing evidence to conclude that substantial undetected fraud has or has not occurred in parapsychology. He considers it possible that Stokes's estimates about fraud may be correct, but those estimates currently must be taken as speculations, not convincing conclusions.

Unfortunately, the uncertainty about the extent of undetected experimenter fraud implies corresponding uncertainty about the validity of the research findings. That was the main point Stokes and I were attempting to make.

My primary purpose in making that point was to bring into focus the need to implement measures to prevent fraud. Both Palmer (personal communication, September 13, 2016) and Roe (2016) indicated that vigilance about the possibility of experimenter fraud is needed and that some measures to address experimenter fraud are appropriate. There appears to be a consensus on this point, although exactly what should be done remains a topic of discussion.

A Methodological Standard for Addressing Experimenter Fraud

Considering all these factors, I believe that the methodological standard of making fraud by one experimenter impossible or very difficult is the optimal practice for research. Experimenter fraud should not be easy and tempting. Implementation of this standard would eliminate the vast majority of

cases of experimenter fraud. I believe that lack of implementation of effective practices to detect and deter experimenter misconduct (fraud and biased methodological practices) invites such behavior and makes undetected cases likely. The research culture in psychology now accepts that methods to prevent questionable research practices are needed. Measures to prevent experimenter fraud should be included in the methodological standards. I consider this standard to be appropriate throughout the behavioral sciences.

As noted above, this standard has long been recommended in parapsychology but has not been implemented in recent decades. Measures to prevent fraud are particularly warranted in parapsychology given the controversial nature of the phenomena, the traditional differences among experimenters in producing effects, and the difficulty in distinguishing between signs of fraud and psi effects. Special experimental designs with extraordinary measures to prevent fraud have also been described (Palmer, 2016; Schmidt, Morris, & Rudolph, 1986; Schmidt & Stapp, 1993); however, these measures are not practical for most research.

My experience has been that it is relatively easy to implement this standard once appropriate research habits have been developed. Measures to prevent fraud are needed for confirmatory research, but are optional for exploratory research by a researcher who plans to conduct one or more confirmatory studies before the findings are published.

Practical recommendations for implementing this standard and implementing other related methodological practices are discussed in Kennedy (2016). One key practice is to make duplicate copies of each component of the data early in the data collection process and handle the copies in a way that would be very difficult or impossible for one experimenter to alter all copies. Ideally, the secure copies will be made before any experimenter has unblinded information that could be used to alter the study results. When that degree of blinding is not possible, two experimenters should be present at any step during the data collection and processing that would allow an experimenter to alter or fabricate data without detection. The experimenters should explicitly and actively have the intention of verifying that intentional or unintentional data alterations do not occur. For automated experiments, documented validation of the software and hardware is needed and if properly done will detect both intentional (fraud) and unintentional problems (Kennedy, 2016).

A healthy, competent research culture will recognize the need to implement such measures as standard procedures. Parapsychological experiments have produced successful results with such measures (Rhine, 1974). Palmer's (2016) concerns about implied accusations of fraud and his speculations about creating a paranoid work environment are not applicable for this type of research culture. All forms of potential research bias should to be openly recognized and addressed. The idea that measures to prevent fraud are implicit accusations of fraud is closely associated with the idea that preregistration and other measures to prevent bias are implicit accusations of intentional bias and that measures to prevent unintentional errors are implicit accusations of incompetence. These types of sensitivities do not have a place in a healthy research culture.

Routine measures to prevent fraud are preferable to the enhanced emphasis on after-the-fact accusations and investigations, which Titus et al. (2008) advocated. Their strategy is based more on happenstance than on a systematic approach. It also asks researchers to take actions that will usually create a major burden for the researchers, including investigations that take a year or longer to complete and that may have inconclusive outcomes. I also expect that reliance on accusations and investigations would produce significant discord in a research environment.

To have reasonable hope of competently evaluating suspicions of fraud in parapsychology, an investigating committee must implement covert detection measures during actual experiments, as was done in the Levy case. That is decidedly not an optimal general strategy for addressing fraud. On the other hand, my experience in research environments with routine measures to prevent fraud has been that the issue of fraud is systematically and effectively addressed with negligible discord. Systematic prevention is vastly preferable to after-the-fact accusations.

Making the raw data available to others for independent analyses is also a useful but secondary strategy for deterring and detecting fraud. Data could be fabricated or altered in a way that does not

leave convincing signs of fraud. As noted above, the cases of detected fraud may be the more extreme, careless frauds. Fraudulent researchers who are more careful may not leave conspicuous signs of fraud. Also, accusations of fraud based on post hoc analyses will too often be circumstantial and irresolvable, particularly in parapsychology.

The possibility of incorrect accusations must be recognized and addressed when accusations are based on statistical analyses. The usual concerns about Type I and Type II errors are applicable, and are enhanced for post hoc analyses. In addition, the statistical methods for screening tests are different than for typical experimental research and should be thoroughly understood if an analyst plans to check or screen a number of studies for statistical evidence of fraud. Anyone making accusations of fraud based solely on statistical analysis would be wise to consult an attorney about the legal liabilities and standards of evidence for libel and slander. The adverse consequences for both sides from the inevitable false accusations that will sometimes occur if statistical methods to detect fraud are widely applied reinforce the point that measures to prevent fraud are much preferable to after-the-fact accusations, including the statistical methods described by Simonsohn (2013). Making the data available to others does not eliminate the need for procedures that prevent fraud.

Efforts to address fraud should avoid any assumptions that the motivations for fraud will be simple and identifiable or that the behavior of those committing fraud will be rational and predictable. The two prominent cases of experimenter fraud in parapsychology cannot be understood in terms of straightforward motivations and rational behavior (Kennedy, 2014). As noted above, there is currently no good evidence for personal characteristics that can be used to predict experimenter fraud. Measures to address fraud should be uniformly applied to everyone.

The most defensible alternative that I can see is to argue that parapsychologists should ignore fraud and focus on developing experiments that can be replicated by any competent researcher. This argument is based on the idea that psi will not be widely accepted in science until virtually anyone can demonstrate the phenomenon. I question whether this idea is true. However, it is clear that parapsychology has not yet achieved the goal of highly replicable results. In order to pursue that goal, the field needs to have experimental results that are worth supporting. That requires studies with good methodology. Funding sources should recognize that measures to prevent fraud and other research biases are good investments and should be a requirement for funding. In addition, I think it is likely that psi is only associated with certain people and under certain conditions. If that is true, good methodology will be essential in making progress in parapsychology.

References

Akers, C. (1984). Methodological criticisms of parapsychology. In S. Krippner & M. L. Carlson (Eds.), *Advances in Parapsychological Research 4* (pp. 112–164). Jefferson, NC: McFarland & Co.

Bierman, D. J., Spottiswoode, J. P., & Bijl, A. (2016). Testing for questionable research practices in a meta-analysis: An example from experimental parapsychology. *PLoS ONE*, 11(5):e0153049, 1–18. Retrieved from http://dx.doi.org/10.1371/journal.pone.0153049

Beloff, J. (1993). *Parapsychology: A concise history.* New York, NY: St. Martin's.

Broad, W., & Wade, N. (1982). *Betrayers of the truth*. New York, NY: Simon and Schuster.

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

Dalton, K., Delanoy, D., Morris, R. L., Radin, D. I., Taylor, R., & Wiseman, R. (1996). Security measures in an automated ganzfeld system. *Journal of Parapsychology*, *60*, 129–147.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4, 1–11. Retrieved from http://dx.doi.org/10.1371/journal.pone.0005738

Gross, C. (2016). Scientific misconduct. *Annual Review of Psychology*, *67*, 693–711

Hansel, C. E. M. (1966). *ESP: A scientific evaluation*. London, UK: MacGibbon & Key.

Hansel, C. E. M. (1980). *ESP and parapsychology: A critical re-evaluation*. Buffalo, NY: Prometheus.

John, L.K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. doi:10.1177/0956797611430953.

Kennedy, J. E. (1975a). Memo to J.B.R. Re: Maze. Retrieved from http://jeksite.org/psi/fmaze.pdf

Kennedy, J .E. (1975b). Summary of Levy egg work. Retrieved from http://jeksite.org/psi/fegg.pdf
Kennedy, J. E. (1975c). Summary of rat implantation work. Retrieved from http://jeksite.org/psi/frat.pdf
Kennedy, J. E. (1980a). Learning to use ESP: Do the calls match the targets or do the targets match the calls? *Journal of the American Society for Psychical Research*, *74*, 191-222. Retrieved from http://jeksite.org/psi/jaspr80a.pdf
Kennedy, J. E. (1980b). Ambiguous data result in ambiguous conclusions: A reply to Charles T. Tart. *Journal of the American Society for Psychical Research*, *74*, 349–356. Retrieved from http://jeksite.org/psi/jaspr80b.pdf
Kennedy, J. E. (2014). Experimenter misconduct in parapsychology: Analysis manipulation and fraud. Retrieved from http://jeksite.org/psi/misconduct.pdf .
Kennedy, J. E. (2016). Is the methodological revolution in psychology over or just beginning? *Journal of Parapsychology, 80*, 156–168. Retrieved from http://jeksite.org/psi/methods_predictions.pdf
Kennedy, J. E. (2017). Notes on a case of scientific fraud in parapsychology. Retrieved from http://jeksite.org/psi/fraud.htm
Palmer, J. (2016). Hansel's ghost: Resurrection of the experimenter fraud hypothesis in parapsychology [Editorial]. *Journal of Parapsychology*, *80*, 5–16.
Retraction Watch (n.d). Retrieved from http://retractionwatch.com/
Rhine, J. B. (1974). Security versus deception in parapsychology. *Journal of Parapsychology*, *38*, 99–121.
Rhine, J. B. (1975). Second report on a case of experimenter fraud. *Journal of Parapsychology*, *39*, 306–325.
Roe, C. A. (2016). The problem of fraud in parapsychology. *Mindfield*, *8*(1), 8–17.
Rogo, D. S. (1985). J. B. Rhine and the Levy scandal. In P. Kurtz (Ed.), *A Skeptic's Handbook of Parapsychology* (pp. 313–326). Buffalo, NY: Prometheus.
Schmidt, H., Morris, R., & Rudolph, L. (1986). Channeling evidence for a psychokinetic effect to independent observers. *Journal of Parapsychology*, *50*, 1–15.
Schmidt, H., & Stapp, H. (1993). PK with prerecorded random events and the effects of preobservation. *Journal of Parapsychology*, *57*, 331–349.
Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science, 24* 1875–1888. doi: 10.1177/0956797613480366
Stokes, D. M. (2015). The case against psi. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st Century* (pp. 42–48). Jefferson, NC: McFarland.
Strobe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670–688. Retrieved from http://pps.sagepub.com/content/7/6/670.full.pdf+html.
Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, *453*, 980–982. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_commentary.pdf

Broomfield, CO, USA
jek@jeksite.org

Abstracts in Other Languages

French

LA FRAUDE DE L'EXPERIMENTATEUR : QUELS SONT LES STANDARDS METHODOLOGIQUES APPROPRIES ?

RESUME : Les discussions sur la fraude de l'expérimentateur en parapsychologie ont manqué de tirer une leçon-clef du cas Lévy. La procédure standard pour gérer la fraude scientifique est une investigation dans l'après-coup (post hoc). On ne peut attendre de ces investigations post hoc qu'elles soient efficaces dans le champ de la parapsychologie car les signes d'une fraude dans les données peuvent être attribués au psi, comme ce fut le cas avec Lévy. En parapsychologie, l'obtention de preuves évidentes d'une fraude nécessite habituellement une détection directe sous couverture de celle-ci au cours de l'expérience, comme dans le cas Lévy. Toutefois, de telles mesures sous couverture réalisées par des collègues ne constituent pas une stratégie très pratique pour traiter la fraude, et elles n'ont pas cours dans les autres domaines de la science. La procédure expérimentale standard pour rendre la fraude par un expérimentateur très

difficile voire impossible a été proposée depuis longtemps en parapsychologie, sans être intégrée au cours des dernières décennies. Elle a toutefois été introduite dans mon expérience de travail dans la recherche médicale réglementée et devrait éliminer la grande majorité des cas de fraude – ceux où l'expérimentateur trouve facile et tentant de manipuler ses données avec très peu de risques d'être détecté. Ce standard fournit un moyen systématique et eficace de s'occuper de la fraude expérimentale et devrait prendre place parmi les nouvelles règles communes de la recherche dans les sciences du comportement.

German

EXPERIMENTATORBETRUG: WAS SIND DIE GEEIGNETEN METHODOLOGISCHEN MASSSTÄBE?

ZUSAMMENFASSUNG: Diskussionen über den Experimentatorbetrug in der Parapsychologie lassen eine Hauptlektion aus dem Fall Levy vermissen. Die Standardvorgehensweise für den Umgang mit wissenschaftlichem Betrug ist eine Untersuchung nach der Tat (post hoc). Man kann nicht erwarten, dass post hoc-Untersuchungen in der Parapsychologie wirksam sind, weil Hinweise auf Betrug in den Daten dem Wirken von Psi zugeschrieben werden können, wie es bei Levy geschah. Ein zwingender Betrugsnachweis in der Parapsychologie erfordert gewöhnlich eine direkte heimliche Aufdeckung des Betrugs, wie sie während eines laufenden Experiments im Falle Levy geschah. Solche heimlichen Maßnahmen seitens von Forscherkollegen stellen jedoch keine praktische Strategie zum Betrugsnachweis dar und sind in anderen Wissenschaftsgebieten nicht üblich. Der Standard, dass durch experimentelle Maßnahmen ein Betrug durch einen Experimentator sehr schwierig oder unmöglich gemacht werden sollte, war lange in der Parapsychologie vertreten, wurde aber in den letzten Jahrzehnten nicht mehr realisiert. Dieser Standard wurde nach meiner Erfahrung in der regulierten medizinischen Forschung eingeführt und sollte die Mehrzahl der Betrugsfälle verhindern – die beginnen, sobald ein Experimentator herausfindet, wie leicht und verlockend die Manipulation oder Fälschung von Daten ist und wie gering die Wahrscheinlichkeit, dabei entdeckt zu werden. Dieser Standard stellt eine systematische und wirksame Art dar, dem Experimentatorbetrug zu begegnen, und sollte Teil der neuen Forschungsstandards in den Verhaltenswissenschaften werden.

Spanish

FRAUDE DEL EXPERIMENTADOR: ¿CUÁLES SON LAS NORMAS METODOLÓGICAS APROPIADAS?

RESUMEN: Las discusiones sobre fraudes en parapsicología perpetrados por el experimentador han hecho caso omiso a una lección clave del caso Levy. El procedimiento estándar para el manejo del fraude científico es una investigación posterior (post hoc). No se puede esperar que las investigaciones post hoc sean efectivas en parapsicología porque los signos de fraude en los datos pueden atribuirse a psi, como sucedió con Levy. Empero, en la parapsicología la evidencia convincente de fraude usualmente requiere la detección encubierta directa de fraude durante un experimento en curso, como en el caso Levy. Sin embargo, estas medidas encubiertas de los colegas no son una estrategia práctica para abordar el fraude y no se practican en otras áreas de la ciencia. La norma de que los procedimientos experimentales deben hacer que el fraude por un experimentador sea muy difícil o imposible se ha propuesto durante mucho tiempo en la parapsicología, pero no se ha aplicado en las últimas décadas. En mi experiencia, esta norma se implementa en la investigación médica regulada y debería eliminar la gran mayoría de los casos de fraude, que comienzan cuando un experimentador encuentra que la manipulación o fabricación de datos es fácil, tentadora, y con muy pocas posibilidades de detección. Tal estándar proporciona una manera sistemática y eficaz de abordar el fraude del experimentador y debe convertirse en parte de las nuevas normas para la investigación en las ciencias del comportamiento.

Journal of Parapsychology, 81(1), 73-86

TECHNICAL NOTE

DISCUSSION ON METHODOLOGY OF WATER CRYSTAL FORMATION AS A DETECTION SYSTEM FOR PSI

By Hideyuki Kokubo, Yasuyuki Nemoto,* and Kimiko Kawano

ABSTRACT: It is often claimed that the form of a water crystal can be changed by human intent although there are few scientific studies that have tested this claim. In 2006, several researchers published the first scientific report which shows positive results for the claim. Recently, the authors have also considered this issue. Unexpectedly, they found two serious weak points in the conventional test procedure for water crystal formation. One of the points relates to the use of judgement, and the other relates to sample size. These points did not seem easy to find, and the authors were concerned that many researchers may fall into the trap laid by these points. In the present paper, the authors discuss these methodological problems of conventional tests, and indicate that photographs of water crystals are not independent parameters and that the numbers of water bottles used are too few to conclude something statistically.

Keywords: water crystal formation, morphological, intent, psi

Liquid water is an excellent solvent which can dissolve a multitude of substances, and it also has important roles in life activities. For example, about 70% of the human body is water, and humans cannot live without water. Water is an important substance for all living creatures on the earth.

In research on ki (qi) and parapsychology, there are studies suggesting that water relates essentially to the nature of anomalous phenomena. For example, Kataoka, Sugiyama, & Matsumoto (1997a, 1997b) reported the results of examining human neutrophils, a kind of leukocyte. After a healer gave his "power" to a water solution (phosphate-buffered saline (PBS), pH 7.4), the solution was given to human neutrophils. It was found that the calcium ion channel on the cell membrane of the neutrophils was opened and phagocytosis was activated. In contrast, the control solution (without the healer's power) did not activate the calcium ion channel nor phagocytosis. Moreover, Sasaki, Sako, and Kobayashi (1993) reported that a qigong master gave his power to water samples and the conductance of the samples changed anomalously during 17 days of measurements. In these studies, healers merely gave their power to water (without any intent to cause a certain phenomenon). These studies suggest that water has a specific property to retain anomalous effects. However, what kinds of mechanisms cause this retention have not been studied yet.

Recently, in chemistry and biosciences, some researchers have reported interesting and strange properties of water. For example, Algara-Siller et al. (2015) reported high-resolution electron microscopy images of water locked between two graphene sheets, an archetypal example of hydrophobic confinement. Their observations showed that the nanoconfined water at room temperature forms "square ice"—a phase having symmetry qualitatively different from the conventional tetrahedral geometry of hydrogen bonding between water molecules. Pollack (2013) found that if liquid water contacted the surface of a hydrophilic material, a special zone of water was formed near the surface. He claimed that it was the fourth phase of water, which was different from solid, liquid, and gas phases. Montagnier et al. (2011) reported that liquid water has an ability to retain DNA information. These studies suggest the possibility to expand consideration of water and to explain ki or psi phenomena scientifically.

On the other hand, as researchers know well, practitioners of healing and qigong often claim that liquid water shows a special responsiveness to ki or psi. For example, the taste of drinks can be changed

if the practitioners give their power to them. Moreover, Emoto (1999) published the photo book *Messages from Water*, claiming that if a person writes phrases such as "love and gratitude" on a paper and attaches the paper to a bottle containing water, water crystals will take special forms. However, these claims are anecdotal and subjective and are insufficient as evidence from the viewpoints of modern science. If researchers want to discuss their claims as scientific hypotheses, experimental studies are needed, for example, to measure the change of measurable parameters (pH, components of foods and drinks, radius of particles, temperature, etc.). Moreover, researchers have to start from a basic hypothesis that anomalous formulation of water crystals can be caused by intent directly, before claiming that anomalous effects also can be caused by written letters. Radin, Hayssen, Emoto, and Kizu (2006) tested water crystal formation in such a basic study and reported that human intent can affect water crystal formation. They tried to replicate the first result in another test (Radin, Lund, Emoto, & Kizu, 2008), but they were unsuccessful. Their attempt stimulated the present authors to look at the anomalous formation of water crystals.

The authors changed the methodology partially and tried to construct a new experimental system which is less ambiguous. The authors tested water crystal formation using a new way of making water crystals. They used two experimental bottles and two control bottles in the same way as in the previous studies. Also they executed two blank tests using two blank-experimental bottles and two blank-control bottles. They took 50 photographs of water crystals for each bottle. A total of 400 photographs were scored by 38 judges from a morphological viewpoint. The authors expected to obtain significant results similar to those of the previous studies. However, unexpectedly, they found two serious weak points in the conventional procedure and concluded that all studies (including their new tests) should be considered as preliminary. One of the points relates to the use of judgement, and the other relates to sample size. These points did not seem easy to find, and the authors were concerned that many researchers may fall into the trap laid by these points. In the present paper, the authors discuss these methodological problems, and they note that photographs of water crystals are not independent parameters and that more experiments are needed to conclude something statistically.

Basic Method

The basic test procedure for assessing water crystal formation, which was adopted by Radin et al. (2006), can be summarized as follows. Human intent is applied to a water-filled bottle for an experiment. Next, paired experimental and control bottles are given to an operator. The operator prepares Petri dishes in which the water crystals are formed, and he or she photographs them the next day.

The work of preparing the dishes is done for each bottle. The operator takes water from a bottle using a syringe; next he puts 1 mL of the water into a Petri dish; then each dish is covered by a lid. In all, 50 sample dishes are prepared from each bottle. The Petri dishes are piled up on a plate and then put into a freezer (-20 degrees Celsius) as shown in Figure 1. The frost in the freezer is removed before the experiment. The temperature of the freezer, which is located in a cold walk-in room, is set to -5° Celsius. After the dishes are in the freezer for about 24 hours, the operator takes their photographs in the low-temperature room using a camera setup.

The operator takes a dish from the freezer, removes the lid, sets it on a microscope, and photographs the dish immediately (Figure 2). The microscope used is a metallurgical microscope (model BX51, Olympus Corp.) with an objective lens (LMPlanFl 20x, Olympus Corp.), and the sample is illuminated from above. Photographs are taken using a digital camera (EOS 6D, Cannon Inc.) with an adapter (NY1S-35, Microscope Network Co., Ltd.) to allow attachment to the ocular tube of the microscope. About 10 s are needed to take the dish from the freezer and take a photo. After taking the photo, the sample is discarded, and the next sample is photographed.

On following days, the operator prepares sample dishes from another bottle and repeats the procedure as given above.

After all samples have been photographed, the photographs are grouped according to which sample bottle the dishes came from. The operator then passes them to the researchers who provide them to the judges.

The photographs are scored by judges on an aesthetic basis. If the ratings of the experimental photographs are higher than those of the control photographs, it is concluded that human intent can influence water crystal formation.

On the surface, this method seems suitable to detect anomalous effects of human intent. However, a few important points must be considered in order to confirm that this method is truly suitable.

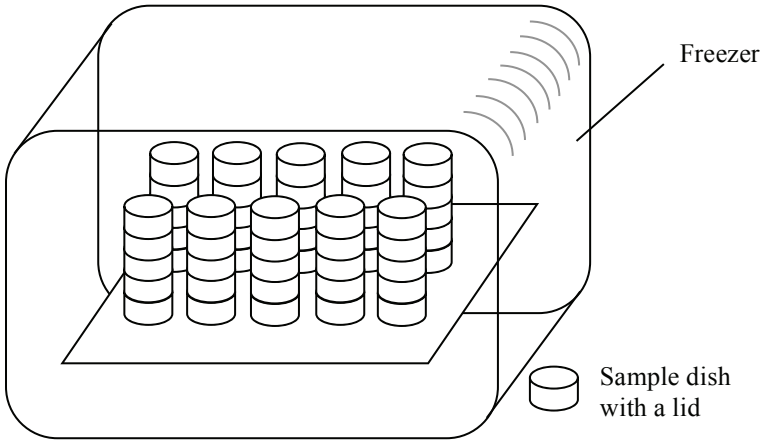


Figure 1. Arrangement of sample dishes during freezing.

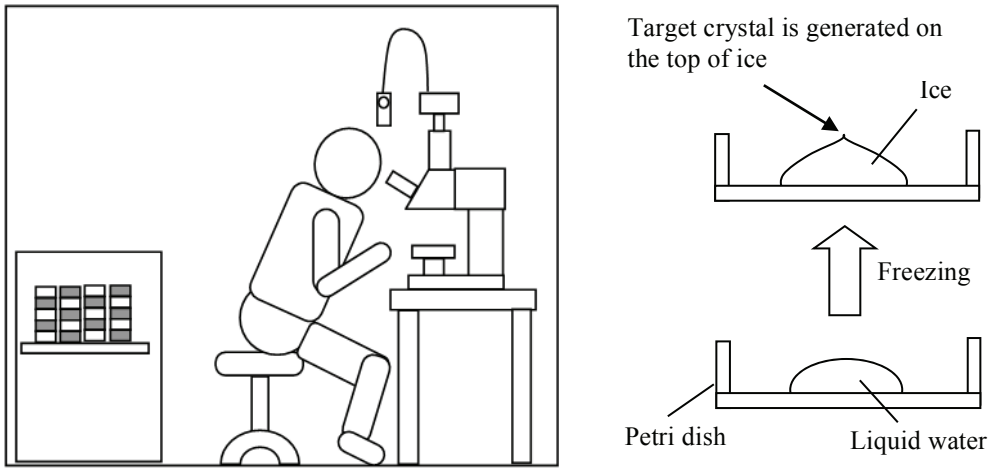


Figure 2. Low-temperature room setup for photography. Several delicate techniques are needed to freeze liquid water to ice having a desirable form, and the I.H.M. laboratory has developed such techniques. Therefore, usually researchers pass their sample water bottles to the I.H.M., where operators photograph the water crystals.

Key Points of Method

Dishes and Phtographs Are Not Independent

One of the key points of this method is that the 50 dishes of water crystals are not independent of one another. Therefore, all dishes are influenced by the same natural fluctuation in the process of freezing and microscopic observation. If the natural fluctuation tilts in a "good" direction, many "good" crystals will be formed. Conversely, many "bad" crystals will be formed if the natural fluctuation tilts in a "bad" direction. The numbers of "good" crystals fluctuate naturally for freezing and observation under the microscope

(Figure 3). Natural fluctuation includes system biases such as humidity, atomospheric pressure, and degree of skill of operators. If an operator fails to freeze water samples for unexpected reasons, all water crystals will become "bad"

Usually, 50 photographs are taken in one experimental period from one sample bottle by one overnight freezing operation. The distribution of rating scores depends on the operation. In other words, the operations or sample bottles are the true independent parameters. It may seem that there are 50 independent photographs, but they are actually derived from one datum.

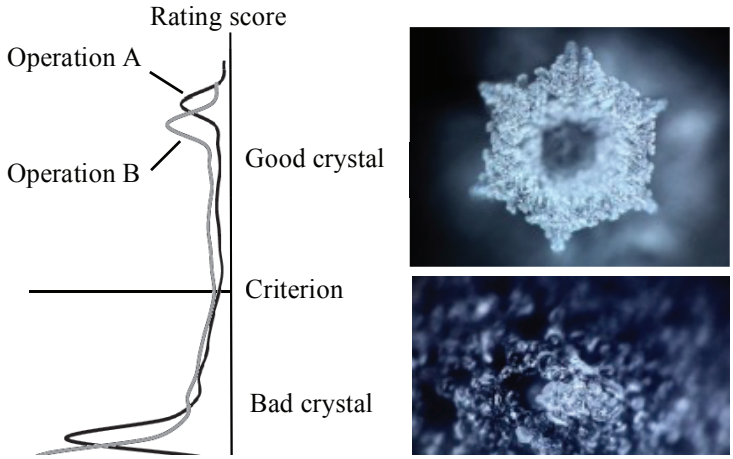


Figure 3. Conceptual distribution of ratings of photographs.

Many Blank Tests Are Needed

Another key point is the number of sample bottles. If researchers want to examine whether an experimental sample bottle shows anomalous effects or not, they should execute many blank tests and estimate the distribution of rating scores of the population of blank tests in advance (Figure 4). The blank test is the experiment to test water crystal formation using water samples without special treatments. Although details of the properties of the population of blank tests are not yet known, the authors expected the *SD* of the blank tests to be relatively large based on the authors' test experiences (details are shown in the next section). Therefore, small numbers of sample bottles are not enough to test for an anomaly statistically. For example, the six bottles in the study by Radin et al. (2008) do not provide useful measures of variability among bottles. They used a hierarchical model that had a factor for each bottle and thus in theory adjusted for differences among bottles. However, only six bottles do not provide a good estimate of variatiability among bottles.

Role of Judges

An additional explanation may be needed for some readers who think that anomalous effects can be detected even from data such as those shown in Figure 3 if many judges are used. It would be correct if we want to know which drawings A or B are liked by people. Many people vote, and then we would be able to discuss a result. If there were many judges, the objectivity of the judgements would increase. However, if we want to know which condition is more effective in forming water crystals, we should compare many data pairs of different conditions. Although a huge number of judges improves the objectivity of the rating scores, the statistical significance between conditions depends on the number of sample bottles. It is impossible to discuss the difference between test conditions statistically if we use only one pair of sample bottles, even if we use many judges.

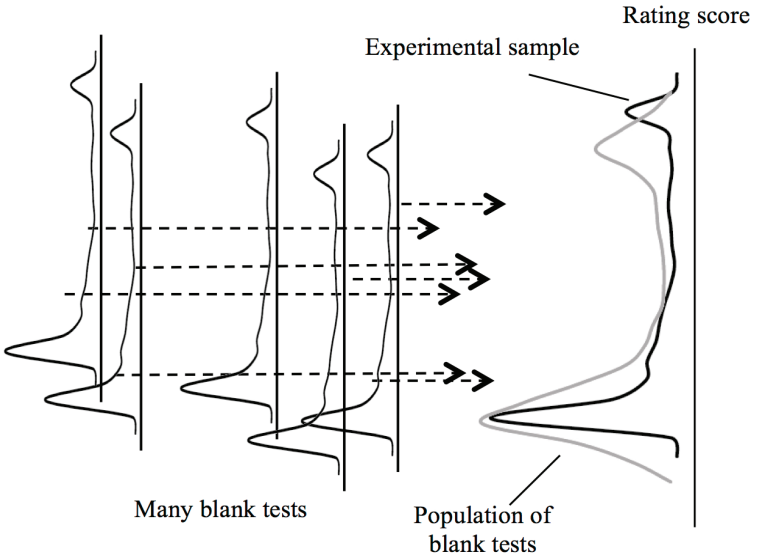


Figure 4. Many blank tests are needed to estimate experimental samples.

Authors' Experiment

In the above sections, the authors discussed methodological problems in conventional tests on water crystal formation. However, those problems seem to be difficult to find if researchers read Radin's papers only. The reason is that many researchers have not experienced such tests and do not have basic knowledge of their properties. In fact, the authors had not succeeded in finding these problems until they did similar tests themselves. In doing their tests, they found the methodological problems mentioned above.

In 2013, H.K., one of the authors, visited the I.H.M. General Institute and had an opportunity to inspect the whole process of photographing water crystals. Following this, H.K. discussed with Y.N., one of the authors, the conventional methodology, and it was decided to change the details of the methodology regarding the seven items described below.

Key Points

1. Research scope. In the present study, the main research question was "Can intent change the form of water crystals?"

2. Type of judgment was changed from aesthetic to morphological. Previously, photographs of water crystals have been rated by whether they are beautiful or not (Radin et al., 2006, 2008). However, aesthetic sensibility differs among individuals. Therefore, it is desirable to adopt a more objective criterion to evaluate photographs. In the present study, to achieve more objective scoring, the judgment of photographs was based on a morphological factor: "How close is the crystal to the hexagon form?"

3. Simple judgment procedure. Judges were university students of classes taught by H.K. Photographs of all the crystals were placed on four sheets (one for each of the paired bottles in each of two conditions), and the photographs on each sheet were arranged according to the order in which they were taken. A total of 24 sets of judgement sheets were prepared. Each set had a different order of the four sheets (4! = 24 sets). Judges rated all photographs by referring to an instruction sheet. Use of judgement sheets is not the best procedure. However, at least, an order effect of bottles is cancelled systematically because each set of judgement sheets is composed of a different order of sheets (bottles).

4. Experimental task based on the judgments. A healer (S017, female) was instructed to intend that the target water forms hexagonal crystals. Therefore, by examining the extent of hexagonal crystal formation, it was easy to determine whether any anomalous effects were caused by her intent, and also to determine whether the intent facilitated or inhibited the hexagonal crystal formation.

5. Equalization of conditions of crystal formation by use of paired experimental and control samples. The target crystal is formed on the top of a block of ice (Figure 2). The core part of a crystal is formed in the freezing process and it grows horizontally to become hexagonal by vapor deposition in a low-temperature environment. Therefore, in the present study the crystal formation process could have possibly been affected by the internal environment of both the freezer and the low-temperature room. If the crystals had been produced on different days, their forms might have been influenced by slight differences among conditions such as temperature, humidity, atmospheric pressure, and adhesion of the frost in the freezer.

In the previous studies using photography, the days on which the photos were taken were different for experimental and control bottles. However, in the present study, the paired experimental and control bottles of water in the main trial were treated simultaneously on a single day. And on another day, the paired experimental and control bottles of water in the blank trial were also treated simultaneously.

Specifically, 50 sample Petri dishes were prepared from each water-filled bottle of a pair of experimental and control trials of the main or blank test, and then the dishes from the bottles were arranged alternatively in the freezer as shown in Figure 5. In this way, minor variations between experimental and control bottles in the circumstances of crystal formation and photography were cancelled. The maximum number of sample dishes was limited to 100 per trial because of the capacity of the freezer and of the time needed to prepare the sample dishes. In addition, during this process, the operator (Y.N.) had not been given information about which bottle was experimental and which pair of bottles was to be used for the main test (double blind condition).

6. Blank condition. In previous studies, not enough consideration was given to fluctuations in the extent of crystal formation. Therefore, in order to estimate the degree of fluctuation, the authors added a blank condition without a healer. The blank trial was executed in the afternoon one day before the corresponding main trial because anomalous remote effects by the healer are more likely if the main and blank tests are done on the same day (Kokubo & Shimizu, 2015).

7. Complete sample of photographs. In the first study by Radin et al. (2006), only 40 of 200 photos were analyzed. This is because the operator passed to researchers only "good" crystal photographs. In their second study (Radin et al., 2008), the operator passed all photographs to researchers. Similarly, in the present study, the authors analyzed all the photographs of all the crystals.

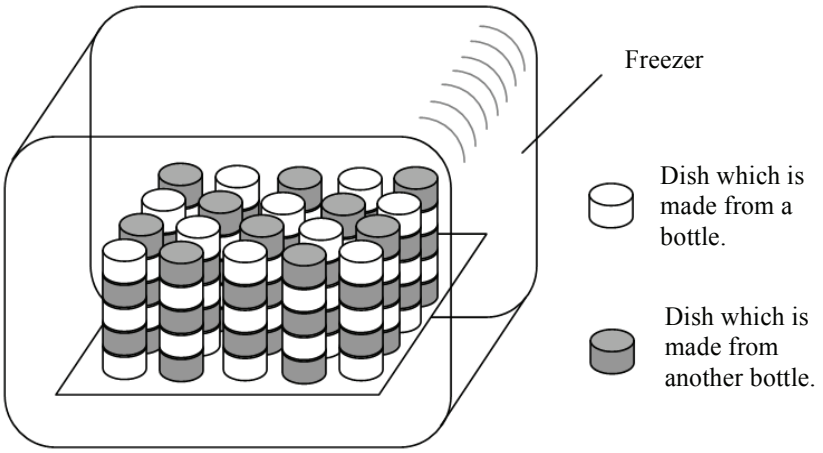


Figure 5. Arrangement of sample dishes in the freezer. All dishes are covered with lids.

Hypothesis

If the healer applies her intent to an experimental water sample with the aim to form hexagon crystals, the extent of hexagonal crystal formation will differ between the experimental and control water samples.

Method

Date and place. This experiment consisted of two sessions, and each session consisted of judgement and a pair of main and blank trials. The first session was done October–November, 2013 and the second session was done March–April, 2014.

At each session, a main trial in which a healer applied her intent to the target water sample was done in the afternoon, while a blank trial was done in the afternoon one day before the day of the main trial at the Institute for Living Body Measurement of the International Research Institute (Chiba, Japan). Preparation of dish samples and photography were done at I.H.M. (Tokyo, Japan) within a few days after the main trial. Photographs of all sample dishes were placed on four sheets (one for each experimental and control bottle in the main and blank conditions), and then the judges gave scores to all the photographs.

Healer. Healer S017 was a 47-year old female who had repeatedly shown large J values ($J > 0.1$) in previous healing studies executed at the International Research Institute (Kokubo, Koyama, Takagi, Kawano, & Yamamoto, 2012; Kokubo et al., 2013a, 2013b; Minami, Usui, & Kokubo, 2014). This study was her first experience in an experiment to form water crystals. Here, J is an index of the magnitude of bio-PK power, and it is defined as a natural logarithm of the ratio of physical measures on experimental and control trials. Physical measures are the intensity of biophotons or gas concentration in the authors' bio-PK experiments. In biophoton experiments, $J = 0.1$ is equal to 0.43 dB. The healer's power can be categorized as: $J < 0.1$, beginner class; $J > 0.1$, middle class; $J > 0.2$, expert class; and $J > 0.3$, psychic class (Kokubo, 2015).

Sample water. The water was purified water (distilled water, 500 mL) initially kept in plastic bottles (Kyoei Pharmaceutical Co., Ltd., Japan; Lot No. 1501).

Double blind. Experimenters were E-1 (H.K.) who did the main and blank trials, coordinated the judging, and analyzed the data; E-2 (K.K., one of the authors) who kept the ID numbers of the sample bottles, and E-3 (Y.N., operator) who prepared the sample dishes for storage in the freezer and took the photographs. After finishing the main trial, E-2 randomly numbered the sample bottles (four bottles: two of them were bottles for a main trial, and others were bottles for a blank trial) and recorded their numbers. E-2 told E-1 only which bottles were paired. Next, E-1 gave this information to E-3 and asked him to prepare the sample Petri dishes, store them in the freezer, and take the photographs.

After the photographs were grouped for each bottle, they were given to E-1. E-1 made 24 sets of sheets of photographs and passed them to the judges. After E-1 made a preliminary analysis of the judgments, E-2 gave the ID numbers to E-1.

The same procedure was repeated at the second session.

Main condition. E-1 chose a pair of bottles at random with a dice, and he used one of them as experimental and the other as control. The healer sat on a chair (Position P in Room A in Figure 6) and tried to generate hexagonal crystals for 30 minutes while holding the water-filled target (experimental) bottle in her hands or setting it on a table. At that time, she used a picture of a hexagonal water crystal on the table as a reference (Figure 7). Room temperature was 23.5–26.5°C and humidity was 22–23%.

The other bottles for this condition were kept at Place C in Room B with the same temperature and humidity.

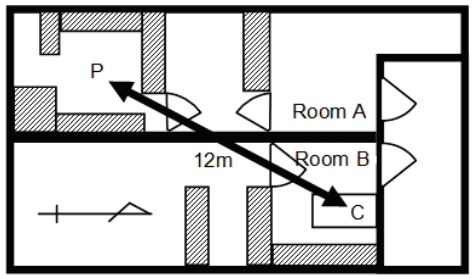
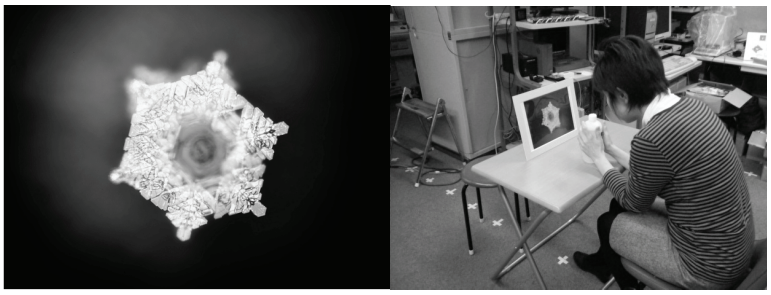


Figure 6. Location of rooms for the experiments. P: Location of target bottle. C: Location of all other bottles. The shaded gray blocks are pieces of furniture.



(a) (b)

Figure 7. (a) A crystal. (b) Main trial with the healer

Blank condition. A blank trial was done one day before each main trial. The blank test procedure was the same as for the main tests except that there was no healer. An experimental bottle was set on a table in Room A for 30 minutes (Figure 8). Room temperature was 25°C and humidity was 26–31%. The other bottles were kept in Room B during the blank tests.

Figure 8. Blank trial without the healer

Randomization. After a main trial, E-2 took the experimental bottle to Room B and then randomly numbered all the bottles for the main and blank conditions. Next, E-1 wrapped all the bottles with packing sheets and put them into a box. E-1 carried the box to E-3 at I.H.M. immediately. E-3 prepared the sample Petri dishes soon after receiving them and stored them in the freezer. At the second session, all bottles of the main and blank trials were treated in the same way.

Growing the crystals and taking their photographs. The water-filled sample bottles (paired experimental and control bottles for the main and blank trials) were given to E-3. He prepared the sample Petri dishes in which water crystals would grow, and he took their photographs the next day. The details of the procedure are the same as described in the section Basic Method, except for equalization of conditions of crystal formation by using paired experimental and control bottles. Photographs were grouped for each sample bottle and then passed to E-1. The same procedure was repeated at the second session.

Judging. E-1 serially numbered the 50 photos of each bottle according to the order in which they were taken. Then he affixed the 50 photos on an A3 size sheet of paper (in total four sheets: two paired experiment/control samples for the main and blank trials), for a total of four bottles per session. These four sheets formed the 24 sets used for judging with the order of the four sheets different for each set.

A reference picture sheet was prepared (Figure 9) for the judges. The judges rated all the crystal photographs according to their conformance to the reference pictures. It took 15–20 min to judge all 200 photos. Judging was done by 20 students for the first session.

At the second session, 200 photos were prepared in the same way as for the first session and judging was done by 18 different students.

None of the students had experience in judging water crystals until this study.

Analysis. The authors counted the numbers of photographs which the majority of judges had classified as A or B. After a preliminary analysis of each session, information on which pair of bottles was for the main trial and which of these bottles was the experimental bottle was given to E-1 and E-3 by E-2.

Results

Distribution of photo ratings. The authors counted the number of photographs which have ratios that were larger than 0.5 and defined the group as A+B photos. Figure 10 shows a distribution of the 400 photos as a function of the proportion of judges who judged a certain photo as A or B. For example, the left-edge bar represents the number of photos (150) which were not rated A or B by any of the judges, and the right-edge bar represents the 30 photos which were rated as A or B by all the judges. As there were 20 judges for the first session but 18 judges for the second session, the X-axis is the proportion of judges who judged a certain photo as A or B.

This distribution can be considered as a distribution of ratings against photos. It is a non-Gaussian distribution having two peaks of numbers of photos at both edges. Many photos were rated as no hexagon.

Descriptive statistics for the A+B photos. Figure 11 shows the number of A+B photos which the majority of judges classified as A or B. Absolute values of the differences between experimental and control bottles in the main trials were larger than absolute values in the blank trials. Especially, in the second main trial, the difference was large. However, the range of ratings of controls in the two main trials (first and second C) and the two blank trials (first and second BE and BC) was large, from a minimum of 4 to a maximum of 17; average was 10.3; *SD* was 5.05 (For all the bottles, the average was 10.8 and *SD* was 4.62). If experimental data of main tests were compared with data for all control and blank bottles, the deviation of the first E was -0.26σ and the deviation of 2nd E was 0.92σ. This suggests that even if human intent can change the formation ratio of "good" crystals, the magnitude of its effect is expected to be equal to or smaller than the magnitude of natural fluctuation.

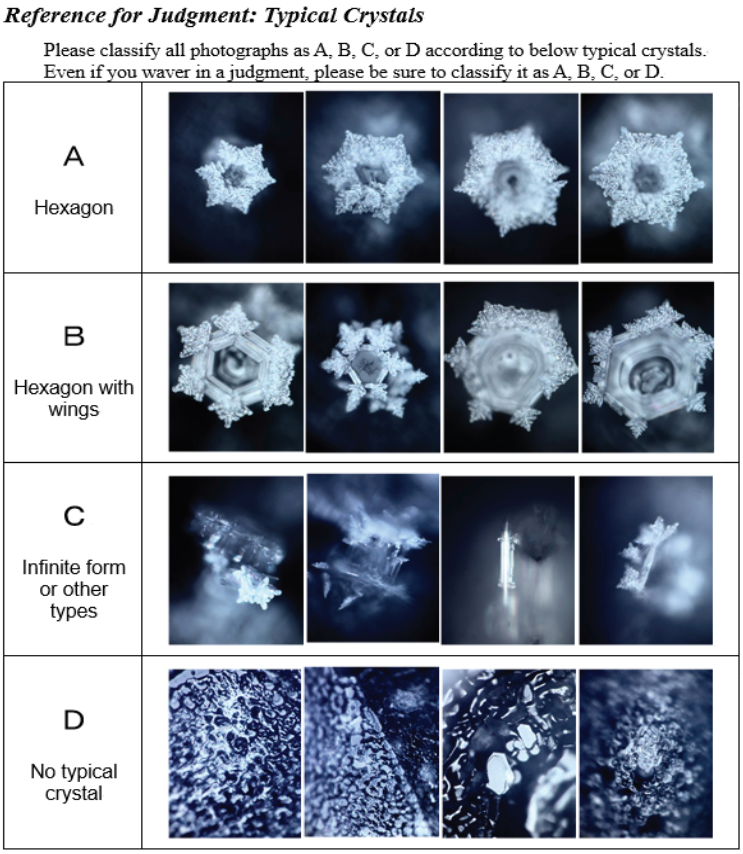


Figure 9. Reference pictures for judging.

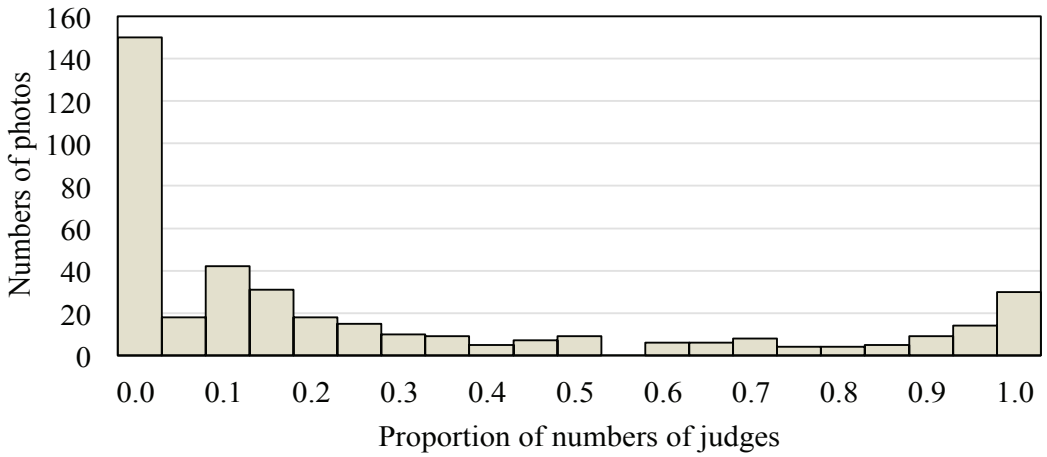


Figure 10. Number of photos rated A or B by a given proportion of judges. The X-axis is a proportion, because the numbers of judges were different for the first and second sessions.

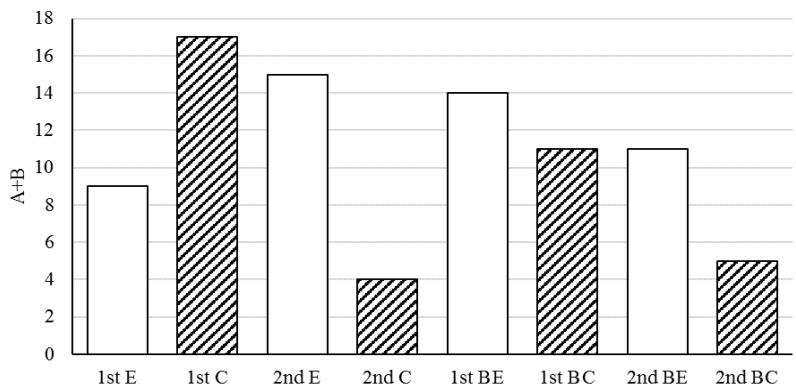


Figure 11. Numbers of A+B photos in the eight conditions. E: Experimental in main test. C: Control in main test. BE: Experimental in blank test. BC: Control in blank test.

Crystal formation as a function of dish position in the pile. Petri dishes were piled atop one another in the freezer. There is a possibility that the formation of water crystals was affected slightly by the position of the dish in the pile, because the lids of the top dishes were exposed directly to the air inside the freezer.

Table 1 shows numbers of photos by position (top and other). The number of A+B photos of top dishes was 11, which is lower than the expected value 17.2. However, the difference of the formation ratio of water crystals was not significant in the present study, $\chi^2(1, N = 400) = 3.55$, $p = .059$, two-tailed.

Table 1
Numbers of A+B Ratings of Dishes at the Top of the Pile Versus Below

position	top	other
A+B	11	75
other	69	245

Although the difference of numbers of photos by position was not significant, the authors consider that the difference may be detected if many sample bottles are used. In fact, the I.H.M. operators claim that

the formation of crystals is often not good in the top dishes based on observation. The reason for this slight difference of numbers of photos by position is still not identified. This is a technical problem in the way of crystals are formed, and it should be improved in the future.

Consistency of judges' ratings. The authors tested the statistical significance of these results by applying the Wilcoxon test to the 20 or 18 pairs of ratings in each trial in Figure 12, and the two-tailed *p*-values at the top of each trial represent the results of these trials. The authors subsequently realized that this test is invalid because, as explained in the Method section, the 20 or 18 pairs of ratings cannot be considered independent. In other words, the proper unit of analysis is not the judge (20 or 18 in each session), but the bottle (2 in each trial). Two bottles are too few to determine the significance of anomalous effects on water crystal formation. However, what is obvious from simply observing the four panels in Figure 10 is that in three of four cases there was high between-judge consistency or reliability of ratings.

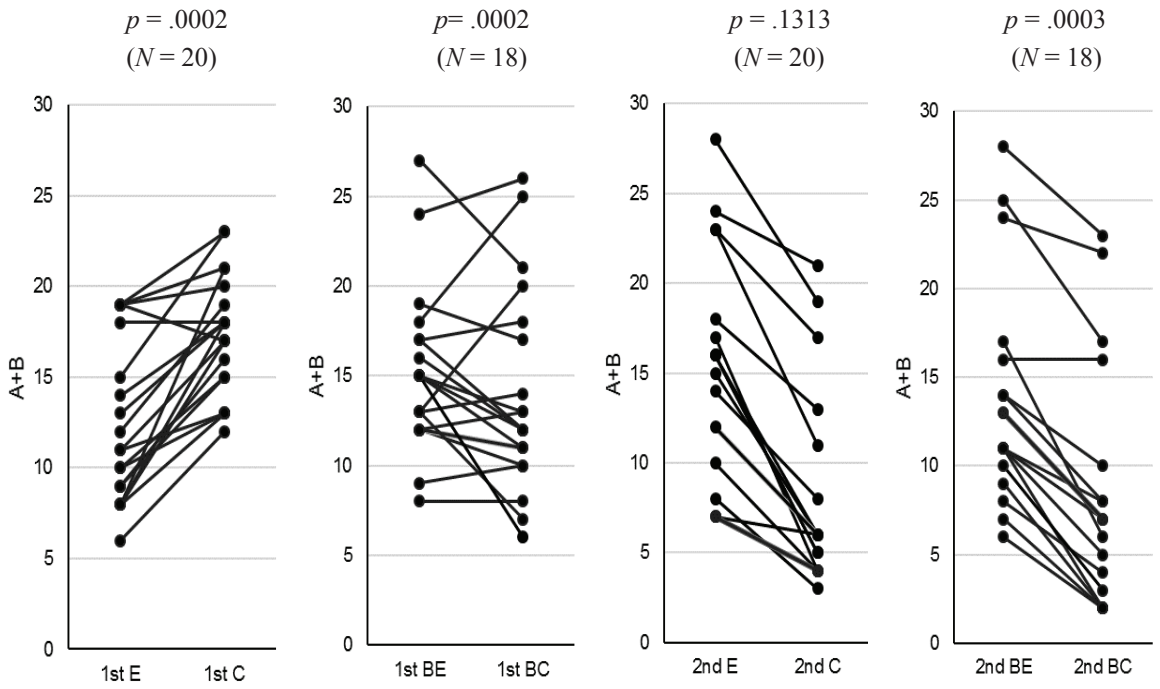


Figure 12. Ratings of the 20 or 18 judges for each pair of experimental and control bottles. E: Experimental in main trial. C: Control in main trial. BE: Experimental in blank trial. BC: Control in blank trial.

Discussion and Conclusion

As shown in Section "Key Points in Method", the authors' tests suggest that researchers (including the authors) had misunderstood the role of judging in conventional tests of water crystal formation, and that two bottles were too few to address anomalous effects. These problems can be relatively easily understood if the researchers recognize them once, or if they try to do the tests the same way that the authors did.

The authors changed the way of making water crystals. Water crystals were made simultaneously for paired experiment and control bottles. In this way, system biases such as temperature, humidity, and the degree of skill in executing the procedure are cancelled, and it is expected that the true effects of human intent can be detected. However, natural fluctuations in the formation ratio of "hexagon" water crystals were greater than expected. In further studies, the existence of natural fluctuation should be taken into account.

The formation ratio of water crystals is considered to be essentially different depending on the experimental conditions. It is difficult to predict the mean chance expectation (MCE) for A+B photos a priori in a given water crystal test. However, the authors consider that this problem can be resolved partially by

use of a standardized index calculated from matched data from experimental and control conditions. For instance, here the authors explain this idea using dB (although they usually use *J*, which is an extension of dB, as an index of the magnitude of anomalous effects in their experiments with cucumber pieces as the bio-sensor; Kokubo, 2015; Takagi et al., 2015). In the present study, the authors used paired experimental and control bottles, and the dishes were placed in a freezer simultaneously (Figure 5). Therefore, the same system biases were applied to the experimental and control dishes. The possible biases are essentially cancelled when paired experimental and control samples are compared. First, the ratio of A+B photos of experimental vs. controls crystals is converted to a dB unit; dB is defined as 10 log(E/C). If there is no special effect and no system bias, the MCE of dB can be expected to equal zero a priori in the blank condition. On the other hand, if human intent causes an anomalous phenomenon, dB will not be zero in the main condition. If main tests are repeated many times, the average of magnitude of the power of the intent can be estimated with confidence intervals. Even if there are some system biases and MCE in the blank condition is different from zero, a true dB can be obtained as a calibrated dB by subtracting the dB from the blank tests from the dB from the main tests (Table 2). Unfortunately, the present sample size (8 bottles) is too small to estimate the magnitude of the power of human intent. However, this analysis method is considered as useful to consider as a way to measure the magnitude of PK power. For example, this method can yield the same values even if the number of photos per bottle is changed from 50 to 100. In further studies, such a standardized index should be considered.

Table 2
dB Analysis of A+B Photos

Pair of bottles	dB	Calibrated dB	*M*	*SD*
1st (blank)	1.047	0		
1st (intent)	-2.762	-3.809	-0.747	4.331
2nd (intent)	5.740	2.316		
2nd (blank)	3.424	0		

Judgements based on a morphological factor can be made easily, even if the judges had not previously rated photos of water crystals. Thus, such a judging strategy is recommended until artificial intelligence (AI) can be applied to the judging process.

Considering the labor involved in an experiment, the present, conventional methods are not necessarily efficient methods. Chauvin (1988) tried to measure the length of the crystal growth of water using capillary tubes. Chauvin's approach differs from the conventional approach in which the form of the water crystal is tested, and the measured parameters are different. However, such an approach should be reconsidered.

The authors do not deny the possibility that crystal formation of water can be affected by intent. However, the sample sizes of previous tests (including the authors' tests) are too small to assess anomalous effects statistically. This possibility should be tested by more experiments, and the magnitude of the natural fluctuation of the system should be estimated with sufficient accuracy in multiple blank tests.

References

Algara-Siller, G., Lehtinen, O., Wang, F.C., Nair, P. R., Kaiser, U., Wu, H. A., Grigorieva, I.V. (2015). Square ice in graphene nanocapillaries. *Nature*, *519*, 443–445.

Chauvin, R. (1988). Modification du degre de congelation de l'eau due probablement a la psychocinese: Auto–observation [Built upon water: psychokinesis and water cooling: A self observation]. *Revue Francaise de Psychotronique*, *1*, 13–19.

Emoto, M. (1999). *Messages from water*. Tokyo, Japan: Hadokyoikusha.

Kataoka, T., Sugiyama, N., & Matsumoto, M. (1997a). Effects of qi-gong vital energy on human neutrophils. *Journal of International Society of Life Information Science*, *15*, 129–137.

Kataoka, T., Sugiyama, N., & Matsumoto, M. (1997b). Effects of qi-gong in relation to the capacity for cancer cell disorder. *Journal of International Society of Life Information Science*, *15*, 458–463.

Kokubo, H. (2015). *Ki or Psi—Anomalous remote effects of mind-body system*. New York, NY: Nova Science.

Kokubo, H., & Shimizu, T. (2015). Pre and post effects in bio-PK experiments. *Journal of International Society of Life Information Science*, *33*, 7–24.

Kokubo, H., Koyama, S., Takagi, O., Kawano, K., & Yamamoto, M. (2012). Bio-PK detectivity of fluorescence measurement method: Early spring test. *Journal of International Society of Life Information Science*, *30*, 208–224.

Kokubo, H., Usui, T., Shimahara, C., Minami, E., Takagi, O., Kawano, K., & Yamamoto, M. (2013a). Response patterns of a bio-sensor in non-contact healing: Analysis by gas and fluorescence measurement methods. *Journal of International Society of Life Information Science*, *31*, 52–60.

Kokubo, H., Usui, T., Shimahara, C., Minami, E., Shimizu, T., Takagi, O., Yamamoto, M. (2013b). Relationships among non-contact healing (bio-PK) ways and response patterns of a bio-sensor— competition and inhibition among healing ways of the chakra-activating type, DNA-rewriting (language-leading) type, and energy-circulation-improving type. *Journal of International Society of Life Information Science*, *31*, 161–181.

Minami, E., Usui, T., & Kokubo, H. (2014). Relationship between age and healing power—a study using a bio-sensor for non-contact healing. *Journal of International Society of Life Information Science*, *32*, 49–55.

Montagnier, L., Aissa, J., Giudice, E. D., Lavallee, C., Tedeschi, A., & Vitiello, G. (2011). DNA waves and water. Retrieved from http://arxiv.org/abs/1012.5166.

Pollack, G. H. (2013). *The fourth phase of water—beyond solid, liquid, and vapor*. Seattle, WA: Ebner & Sons.

Radin, D., Hayssen, G., Emoto, M., & Kizu, T. (2006). Double-blind test of the effects of distant intention on water crystal formation. *Explore*, *2*, 408–411.

Radin, D., Lund, N., Emoto, M., & Kizu, T. (2008). Effects of distant intention on water crystal formation: A triple-blind replication. *Journal of Scientific Exploration*, *22*, 481–493.

Sasaki, S., Sako, Y., & Kobayashi, T. (1993). Characteristics of qi observed in the changes of qi-water conductivity. *Journal of Mind-Body Science*, *2*, 1–7. [In Japanese]

Takagi, O., Sakamoto, M., Yoichi, H., Kokubo, H., Kawano, K., & Yamamoto, M. (2015). Discovery of an anomlous non-contact effect with a pyramidal structure. *International Journal of Sciences*, *4*, 42–51. DOI: 10.18483/ijSci.714

International Research Institute
40A, Yuuki Bldg., Sonno 1108-2
Inage, Chiba 263-0051, Japan
kokubo@a-iri.org

**Yasuyuki Nemoto*
I.H.M. General Institute
2F, NS Bldg., Higashi-Nihonbashi 2-6-1
Chuo, Tokyo 103-0004, Japan
yasuyuki.nemoto@hado.com

Abstracts in Other Languages

French

NOTE TECHNIQUE

DISCUSSION SUR LA METHODOLOGIE DE LA FORMATION DE CRISTAL D'EAU COMME SYSTEME DE DETECTION DU PSI

RESUME : Il est souvent affirmé que la formation d'un cristal d'eau peut être modifiée par l'intention

humaine bien que peu d'études scientifiques ont mis à l'épreuve cette affirmation. En 2006, plusieurs chercheurs ont publié le premier compte rendu scientifique qui montrait des résultats positifs en faveur de cette affirmation. Récemment, les auteurs ont également étudié cette question. De façon inattendue, ils ont découvert deux faiblesses méthodologiques graves dans la procédure conventionnelle de test de la formation de cristal d'eau. L'une de ces faiblesses est relative à l'utilisation d'un jugement, l'autre à la taille d'échantillon. Ces points ne sont pas faciles à détecter, et les auteurs craignent que d'autres chercheurs puissent tomber dans les mêmes chausses-trapes. Dans cet article, les auteurs discutent ces problèmes méthodologiques des tests conventionnels, et indiquent que les photographies de cristaux d'eau ne sont pas des paramètres indépendants et que le nombre de bouteilles d'eau utilisé est trop faible pour permettre de conclure sur le plan statistique.

German

TECHNISCHE MITTEILUNG

DISKUSSION ZUR METHODOLOGIE DER WASSERKRISTALLBILDUNG ZUM PSI-NACHWEIS

ZUSAMMENFASSUNG: Häufig wird behauptet, die Form eines Wasserkristalls ließe sich absichtlich durch Menschen beeinflussen, obwohl es nur wenige wissenschaftliche Studien gibt, die diese Behauptung überprüft haben. 2006 haben mehrere Forscher die erste wissenschaftliche Arbeit dazu veröffentlicht, die positive Ergebnisse erbrachte. Die Autoren haben sich kürzlich über dieses Thema nochmals Gedanken gemacht. Sie stießen dabei wider Erwarten auf zwei gravierende Schwachstellen beim konventionellen Testverfahren zur Wasserkristallbildung. Einer der Punkte betrifft die Art der Einschätzung, der andere die Größe der Stichprobe. Diese Punkte waren nicht leicht zu erschließen, und die Verfasser befürchten, dass viele Forscher durch diese Punkte in die Irre geführt werden könnten. In der vorliegenden Arbeit diskutieren die Autoren diese methodologischen Probleme der konventionellen Tests und weisen darauf hin, dass die Fotografien der Wasserkristalle nicht unabhängig voneinander sind und dass die Anzahl der verwendeten Wasserflaschen zu gering ist, um statistische Schlussfolgerungen zu ziehen.

Spanish

NOTA TÉCNICA

UNA DISCUSIÓN SOBRE LA METODOLOGÍA DE LA FORMACIÓN DEL CRISTAL DE AGUA COMO SISTEMA DE DETECCIÓN PARA PSI

RESUMEN: Se afirma a menudo que la intención humana puede cambiar la forma de un cristal de agua aunque pocos estudios científicos han puesto a prueba esta afirmación. En 2006, varios investigadores publicaron el primer informe científico que apoyó la hipótesis. Recientemente, los autores de este trabajo también consideraron esta cuestión. Inesperadamente, encontraron dos vulnerabilidades serias en el procedimiento convencional para evaluar la formación de cristal de agua. Uno de los problemas se refiere a la evaluación y el otro al tamaño de la muestra. Estos puntos no son fáciles de detectar y los autores consideran que muchos investigadores pueden dejarse engañar por estos puntos. En este trabajo, los autores discuten estos problemas metodológicos de las pruebas convencionales e indican que las fotografías de los cristales de agua no son parámetros independientes y que se usaron demasiadas pocas botellas de agua para concluir algo estadísticamente.

Journal of Parapsychology, 81(1), 87-95

BOOK REVIEWS

THE ESSENTIAL GUIDE TO REMOTE VIEWING by Paul H. Smith. Las Vegas, NV: Intentional Press, 2015. Pp x + 280. $17.95 (paperback). ISBN 9-7819-3881-5010.

Much information has been written about the many facets of remote viewing. (RV). One can find such information on the Internet and in books and magazines. Paul Smith has been able to succeed in condensing the many facets of remote viewing into this relatively short how-to-remote viewing guide.

Paul discusses how remote viewing was used in military intelligence gathering, how it has been used in police investigations and finding missing people, and how it is being used in financial investing. The book is not just a presentation of theory and hands-on remote viewing, but it also discusses actual cases where remote viewing successfully helped to provide information that actually solved problems and answered questions.

Paul's writing is easily understood and to the point. It is like he is talking to you. He presents remote viewing in a logical manner. This guide explains exactly what remote viewing is and what it is not. Paul sufficiently explains the different types or methodologies of remote viewing as well as the protocol involved in a remote viewing session. The protocol is what makes remote viewing remote viewing.

Paul starts by discussing the most popular of remote viewing styles, which is coordinate remote viewing (CRV) brought into existence in the late 1970s and the early 1980s by Stanford Research International (SRI) researchers and Ingo Swann, a SRI remote viewer. CRV is made up of six stages that are well explained in the guide book by discussing how the viewer can gain access to the target site in stages and what one can expect to experience in each stage. Paul goes on to explain how remote viewing evolved over the years by the efforts of people such as his former colleagues Lynn Buchanan, David Morehouse, and Ed Dames. He mentions Courtney Brown and Glenn Wheaton, two individuals outside the government system who are currently teaching their own remote viewing methodologies.

Paul gives much discussion to associative remote viewing (ARV) and explains how it is not a remote viewing method. Instead, it is a way of tasking a remote viewer to try to receive a certain kind of information. Any remote viewing methodology can be use used for ARV.

Overall, despite what remote viewing methodology one chooses to use, Paul offers the tools and the structure on how one can successfully remote view. His guide book teaches that given some time and effort and a focused intention, anyone can remote view**.** In the guidebook, Paul discusses how ARV is used to make decisions about the future. An example of when ARV is used is trying to determine whether a stock will go up or down in value at a given time (the closing time of the stock market). Another example of ARV is determining what sports team will win an event. Please notice that in ARV a binary outcome is involved.

ARV works by associating a descriptive target with each outcome. The target associated with what will be the correct outcome will be viewed by the remote viewer before the event takes place. After the event has occurred, the remote viewer is then presented with the correct target as feedback. Paul explains that ARV is great to use because if one asks the viewer outright who will win a sports game the response would be no better than chance. There will be only a 50/50 chance that he or she will be right. In other words, it would essentially be a guess.

Paul tries to explain remote viewing scientifically as best he can. He does, however, write about the evidence of remote viewing. He first discusses the difference between anecdotal evidence and statistical evidence. Anecdotal evidence tends to have limited value. We have to rely on what people are saying and hope that it is true. The significance of statistics and statistical data immensely helps give credibility to the field of remote viewing. He cites how Jessica Utts, a statistics professor, made a strong case for the legitimacy of remote viewing.

Besides learning the basics of remote viewing, the reader is introduced to the pioneers in the field as

well as meeting some newcomers. It has been well documented that Russell Targ and Hal Putoff coined the word "remote viewing" when they were conducting parapsychology at Stanford Research Institute. I found the forword in the book by Sally Rhine Feather, daughter of J. B. and Louisa Rhine, interesting. It gave insight to the private lives of two famous psychic researchers. Paul's story about his son as a researcher in associative remote viewing at Colorado University is interesting and amusing. Paul shows us how new remote viewers took to the field.

I think the best and some of the more exciting information in the book is about some of the astonishing successes of remote viewing not only by professional viewers but also by first time viewers. Chapter 4 in the book discusses and shows accurate drawings and sketches of the Soviet Typhoon-class submarine remote-viewed by Joseph McMoneagle and Hartleigh Trent in early 1979.

Before I close my remarks, I would like to say that I think the pictures of individuals, targets, and even the old government building that once housed the government psychic program, as well as the remote viewing sketches, add great interest to the book. They help tell the whole story of remote viewing. As Paul says in the book, "But knowing the history isn't necessary to learn remote viewing" (p. 251). This, of course, is true. I cannot help but think that people would be somewhat amazed of how remote viewing came about and follow its evolvement throughout the decades.

In closing, I would like to say that *The Essential Guide to Remote Viewing* is a very informative and insightful book on remote viewing. I would recommend it to the lay person to read. In fact, I would recommend it to anyone to read who is interested or curious about remote viewing.

Angela Ford

11213 Lakeview Drive
Dunkirk, MD 20754
fordangela04@gmail.com

KI OR PSI—ANOMALOUS REMOTE EFFECTS OF MIND-BODY SYSTEM: BIOPHYSICAL APPROACH TO UNKNOWN POWER by Hideyuki Kokubo. New York, NY: Nova Science Publishers, 2015. Pp. vii + 196. $190 (Hardback). ISBN 978-1-63482-954-0.

The concept of bio-energy has existed in different cultures for centuries. In China, this force is called *chi or qi*. In India, it is called *prana*. In Japan, it is called *ki*. Despite the different names, each term describes an observable, active force that represents the difference between a living, active system and an inactive and stagnant system. Inanimate objects are "brought to life" by forces such as electricity, air flow, or any number of other activating factors, but ki is a force considered to interact with biological systems and animate flesh and biological materials that would otherwise be still and essentially dead.

In Eastern philosophies and modern Japan, ki not only is described as providing animation, but also is recognized as a force that can be manipulated, projected, and applied by a *ki master* or energy healing practitioner. Though there may be a tendency to relate ki to descriptions of the soul or spirit, typical definitions of soul or spirit do not describe a force that can be projected and consciously applied by a trained master. In contrast, ki is most often recognized in the directed application of force or energy by a master of specific disciplines or artistic practices.

Although the main title of this book, *Ki or Psi*, may give the impression that the author will present a comparison between the concepts of ki and psi activities, ki is actually described in this book as a form of psi energy, specifically a force that most parapsychologists would call bio-PK. So rather than posing a question, the title is actually proposing an equivalency between the terms ki and psi and presenting these forces as activating factors in the demonstration of nonlocal healing or remote effects on biological systems. The value of this equivalency will have to be determined by the reader as the concept of ki is described in detail based on the effects observed in the laboratory studies presented in this book.

Ki or Psi is a complex book that helps the reader to understand the basic concepts of ki from the perspective of Japanese culture, explores how ki may be used in energy healing activities, discusses specific research studies conducted by the author to attempt to detect ki, provides guidance for researchers who wish to perform similar studies, and, finally, briefly summarizes the history of research performed in Japan relating to ki, energy healing, and other psi effects.

The book was originally written in Japanese and translated by its author into English, and though the translation is awkward at times, overall, the presentation of the material is well represented in English. That is not to say that the topic is clear and easy to understand. This book is extremely technical, including explorations into the physics of electronic noise, the chemistry of measuring gas emissions, the biology and structure of plants (specifically cucumbers), the development of statistical analysis methods, parapsychological concepts, and a historical review of existing research. Due to the technical nature and the range of topics that are covered in this book, it can be difficult to follow. Though it is under 200 pages, each page is packed with information and concepts that require careful reading in order to progress to subsequent sections. Having said this, some of the sections are reasonably self-contained and can be used as independent guides if the reader has a basic understanding of the concepts of ki and experimental methodologies.

This book consists of five chapters, each filled with details and information concerning ki and research performed around ki and energy healing.

- Chapter 1: Concept of Ki and Bio-Sensors
- Chapter 2: Research on Ki Using Bio-Sensors
- Chapter 3: How to use Cucumber As a Bio-Sensor
- Chapter 4: Biophoton and Florescence Measurement Methods
- Chapter 5: Brief History of Japanese Studies

Chapter 1: Concept of Ki and Bio-Sensors

Ki is a topic that is seldom explained and rarely examined in detail. In this chapter the author takes the time to explore historical and cultural perspectives on ki and provide definitions in terms of the most modern research and common societal terms. He discusses the linguistic origin of the term and how it is traditionally perceived in medical treatments and demonstrations of physical effects that result from the application of ki.

Bio-sensors are also explored beginning with descriptions of physical sensors, physiological sensors, and the correlated activities between human intention and biological entities such as cells, animals, and plants. His descriptions also include how quantum processes are used as bio-sensors in the random event generators deployed with the Global Consciousness Project. Finally, the author introduces the cucumber as an instrument to measure bio-energy, including a detailed evaluation of the specific variety of cucumber that he considers most appropriate for use as bio-sensors.

This chapter includes detailed instructions for the preparation of the cucumbers for experimental purposes, including the appropriate width of cucumber slices, whether slices should come from the flower or vine side of the vegetable, and how samples should be arranged for optimal comparison. There are extensive references in this chapter related to the species of cucumbers and other references to the Global Consciousness Project, but certain topics regarding the preparation of cucumbers and their treatment appear to have originated in the author's personal experiments, since there are nearly no references to support these proposals.

The chapter ends with the description of a quantitative measurement of the magnitude of ki energy called *J*. *J* is calculated by taking the logarithm of the ratio between data collected from the experimental and control samples. It also includes a coefficient based on the sensitivity of the bio-sensor being used. As a result of numerous experiments with different healing participants, the author defines *J* values between 0.1 and 0.3+ as identifying beginner healers, mid-level healers, experienced healers, and psychic healers, respectively.

Chapter 2: Research on Ki Using Bio-Sensors

In studies of master healers and healers in training, the author and his colleagues examined the effects of various healing activities such as reiki, johrei, qigong, and others. Using cucumbers as bio-sensors, the studies focused on exploring how well the cucumbers could register human healing intentions and the nature of the effects.

To justify the use of cucumbers as bio-sensors, the author presents an explanation that I will attempt to summarize before we begin to explore the details of his studies. The explanation begins with a claim that cucumbers are biological matter that would respond to ki if it were to be applied during a period of biological distress. Of course, many other biological entities may also respond to the application of ki, but cucumbers have a much simpler structure than animals and no belief system to complicate the healing process. Also, the characteristics of cucumbers are well understood and documented by biologists and horticulturalists. By cutting the cucumbers, the researcher introduces an injury to the vegetable that activates the vegetable's attempt to recover from the injury. This attempted healing activity produces an excess of biophotons that increases as greater healing activity occurs. By measuring the amount of biophotons from a cut cucumber slice, the extent of healing activity can be measured, and by comparing cucumbers that are treated by energy healers with cucumbers that are in a control group, the extent of healing provided by the energy healers can be quantified.

Two basic claims are not clearly established in this description. First, why would a cucumber that has been removed from a living, growing plant continue to respond to an injury as if it were still connected to the plant? It is clear that plants attempt to recover from broken branches or torn leaves, but after the cucumber is removed from the growing plant, it is not clear why this healing should continue. Second, if the cucumber were to demonstrate a healing type reaction after it is cut, what evidence demonstrates that it would respond to ki as a healing factor or catalyst? Neither of these questions is clearly answered or demonstrated by referenced studies. Nonetheless, the experiments that follow are valuable in their demonstration that the effects of human healing intention provoke certain measurable factors in the cucumbers.

The first experiments were performed to see if heat or shadows from a participant's hands could have an effect on the cucumbers being used as bio-sensors. Next, magnetism was tested as an additional mediating factor. In each case, the cucumbers were prepared using a specific procedure (described in Chapter 1) and the experimental and control samples were each monitored for the emission of biophotons (light in the range of 280–650 nm) over a period of approximately 18 hours. Measurements were taken continuously through this period, and cumulated each hour. *J* values (see Chapter 1 summary) were calculated for each condition and no significant variance was reported for heat, shadow, or magnetism when no healers were involved in the process.

A number of healers in the studies did achieve significant results, which led to an investigation of character traits or behaviors that might affect the expression of ki in these specific experimental procedures. Neither specific character traits nor age of the healer seemed to have a direct correlation with the *J* values recorded in these studies, so further studies were conducted to differentiate the approach taken to healing by each healer. Their activities and beliefs were surveyed along with a number of other factors that might affect the healing effect. Though there were some relationships between differences in the techniques and apparent effectiveness in specific situations, there was no indication that any technique or approach was superior or should be a preferred method for research purposes.

One of the most intriguing reports in this chapter is a report of the shape and size of the healing fields generated by the most effective healers. The studies involved the creation of a pattern of cucumbers, as bio-sensors, in different locations and different distances from the healer. The healer attempted to produce healing effects on all of the cucumbers, and the cucumbers were measured to see which demonstrated the highest levels of biophotons. The resulting data indicated a wavelike pattern of healing effect that were high and low in alternating distances from the healer, as if peaks and valleys of a wave were being captured by the cucumbers, demonstrating an uneven pattern of healing effects (Kokubo, et. al., 2010). This pattern of effects is reminiscent of the wave-like patterns of activities described by Joines in his examination of poltergeist activity (Joines, 1975; Roll, 2003) which provides an interesting comparison between the effects

observed in the unintentional PK activity of poltergeists and the intentional healing activities demonstrated in these studies. I found myself wishing for a more detailed examination of this topic by the author.

Chapter 3: How to Use Cucumber as a Bio-Sensor

The third and fourth chapters describe original methods developed by the author as a means of detecting Ki using cucumbers as sensors. Chapter three goes into details about the selection of the cucumber species, the optimal time for harvest and storage, and the preparation of cucumber slices for inclusion in the experimental and control samples. Evaluation methods are described in detail, including specific examples of how to calculate and interpret the *J* values.

Important considerations mentioned by the author include the proper approach to inform participants of the purpose of the study, methods to protect research measurements from electrical noise and interference, and a method to evaluate specifically when "power" is being generated by a healer versus the healer imagining they are creating healing energy. As is the case in other sections of this book, the readers must determine whether they accept the word of the author based on the stated experimental findings because most of these factors are not referenced to published articles. Though such articles may exist, the references are absent.

Specific research methodology is outlined, including timing for each session and the isolation of control samples. The methods described would be valuable for researchers who wish to replicate these studies or design their own studies based on the use of cucumbers as bio-sensors.

The final portion of the chapter is an explanation of the gas discharge method that is focused on the detection of hexanol and measurement of the differences between the emissions of the experimental and control samples. The equipment, methods, timing, preparation, and cleanup are all described, to enable a researcher to create a similar setup. There is no clear explanation that describes why hexanol is representative of healing activities in the cucumbers, but an increased presence of hexanol was recognized when the healers were intentionally attempting to apply ki to the cucumbers.

Chapter 4: Biophoton and Fluorescence Measurement Methods

This is a short chapter explaining the methods used to measure biophotons as well as explaining how measurements of fluorescent light may determine if healing is occurring in cucumbers. Biophotons are described as ultraviolet light in the range of 280–650 nm and higher levels of biophotons are identified as an indication of increased healing activity, presumably as the result of ki or intentional healing activity. Biophotons are measured using an image intensifier system mounted on a dark box that is climate controlled and designed to contain the cucumber slices during post-healing measurements.

Florescent light measurements had been used in previous studies in which an increase of green-yellow light was an indication of ki activity (Kokubo & Yamamoto, 2012). Kokubo stated that the current studies were determined to be insufficient to recommend using florescent light emissions to measure cucumbers as bio-sensors for ki.

Chapter 5: Brief History of Japanese Studies

As the chapter title indicates, this is an inventory of Japanese and some Chinese studies on psi phenomena including studies of ki and ki masters. The chapter begins with studies from 1946–1973, progresses through studies examining Uri Geller in the 1970s, and continues up to current-day activities. The extensive bibliography in this chapter is very useful for researchers interested in learning more about psi research in Japan and China.

Final Thoughts

This publication is densely constructed with detailed technical information on numerous scientific and cultural topics. In the descriptions of his experimental approach and justification for his research, I

found that I was finally able to understand why Kokubo chose cucumbers as bio-sensors and how he justified his interpretation of biophoton emissions as representative of healing activity (See summary of Chapter 2). If only for this information, this publication provides significant insight to help justify the extensive line of research this author initiated to explore the nature of ki. Though the bio-sensor measurements are essentially measurements of correlation between biological matter and human intention, these studies provide insight into the activities of self-identified energy healers and how biological matter responds to their efforts. As the author mentioned in his final remarks, "If researchers want to reveal the mechanisms of ki or psi, they need several laws which are described by equations. The laws should be induced from experimental data" (p. 183).

This book attempts to formalize the experimental processes necessary to generate equations that can be used to create laws that describe the behavior of ki. Kokubo presents excellent research methods designed to reveal the inner nature of ki, but additional questions arise from these studies. Will this method work with other, more traditional sensors? Why are these effects interpreted as healing in these sensors, and how has this effect been associated with ki?

If these studies continue to be refined and repeated with more traditional bio-sensing technology and if a larger sample of healers is surveyed for the studies, eventually the laws that govern ki or psi may be more clearly defined.

References

Joines, W. (1975). A wave theory of psi energy [Abstract]. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.) *Research in parapsychology 1974* (pp. 147–149). Metuchen, NJ: Scarecrow Press

Kokubo, H., Takagi, O., Koyama, S., & Yamamoto, M. (2010). Spatial distribution of potential of controlled healing power—Exploratory measurement using cucumber as a bio-sensor. *Journal of International Society of Life Information Science*, *28*, 236–249.

Kokubo, H., & Yamamoto, M. (2012). Fluorescence measurement method for non-contact healing power. *Journal of International Society of Life Information Science*, *30*, 41–48.

Roll, W. G. (2003). Poltergeists, electromagnetism, and consciousness. *Journal of Scientific Exploration*, *17*, 75 – 86.

John G. Kruth

Rhine Research Center
2741 Campus Walk Avenue
Building 500
Durham, NC 27705, USA
john.kruth@rhine.org

PSYCHIC PHENOMENA AND THE BRAIN: EXPLORING THE NEUROPSYCHOLOGY OF PSI by Bryan J. Williams. Gladesville, NSW, Australia: Australian Institute of Parapsychological Research, Inc., 2015. Pp. xii + 135. ISBN 978-0-9870772-2-6.

Parapsychology is the study of conscious experiences including extrasensory perception (ESP: telepathy, clairvoyance, and precognition) and psychokinesis (PK: interaction of mind and physical matter) that are described as paranormal or anomalous because they appear to transcend our current understanding of physical laws. Neuroscience is the study of the structure and function of the brain and nervous system based on the belief that all human experience has its origins in brain activity. It would appear that these two scientific domains have little in common and cannot even speak to each other. Indeed, many neuroscientists over the years have expressed the opinion that ESP and PK cannot exist, because there is no plausible explanation for them in terms of brain activity. However, the field of parapsychology has had a long-standing fascination with neuroscientific research, looking for the fundamental explanations of psi in terms of the physical mechanisms of brain function.

This monograph, published by the Australian Institute of Parapsychological Research, brings to-

gether decades of research linking psychic phenomena and the brain. The author, Bryan J. Williams, has written on this topic in both academic review articles and book chapters. He was a research student of William G. Roll until his death in 2012 and is currently affiliated with the Psychical Research Foundation in Carrollton, Texas.

Williams' monograph serves as an introduction to this area of inquiry, well-suited to the layman or to those students who are entering into the field of parapsychology. It provides the essential information about the history of parapsychology research and the methods of neuroscientific investigations of brain activity. This material is a necessary prerequisite for an understanding of the discussion of the research studies that comprise most of the text. The monograph provides access to the scientific literature with hundreds of reference citations to the original research, which enables deeper exploration of the areas of greatest interest to the reader. The text is clear and free of jargon, which makes it accessible to those who are not already familiar with the material.

The preface and first chapter lay out the basic issue. The mechanisms of common sensory perception (sight, hearing, touch) have been thoroughly investigated. Specific sensory end-organs receive physical stimulation signals that travel through the nervous system and are processed, filtered, and integrated into perceptions within conscious subjective experience. Movement and the motor system are governed in a similar way by other components of the nervous system. However, there are instances in which people claim to perceive or to know without any apparent stimulation of the physical sense organs. As well, there are instances when individuals appear to influence events in the physical world without physical contact and the use of the motor system. These are the realm of parapsychology, the subjects of rigorous study for more than a century. Accumulated evidence supports the existence of these phenomena, in both the real world and the laboratory, which raises the questions of how they can possibly occur and whether psi phenomena can be accounted for by actions and functions of the brain and nervous system?

For many scientists, who believe that every perception, thought, emotion, and action can be completely explained by the chemical and electrical activity of the brain, the idea that perception or action can take place through mechanisms beyond brain activity leads to a flat-out rejection of the very possibility of psi phenomena, despite any evidence. Williams suggests that an examination of the research on neuropsychological correlates of psi activity could provide a bridge between these two opposed areas of scientific inquiry. For Williams, the research might show that psi phenomena are indeed connected to the brain, or at least are not far removed from it. Such a bridge could advance both neuroscience and parapsychology.

The next chapter presents essential background on parapsychology and on neuroscience. Williams describes the different psi phenomena (telepathy, clairvoyance, precognition, and psychokinesis) and provides an historical summary of research on each. Although this material will be familiar to those in the field, this introduction will be of tremendous value to those new to the field, who are the intended audience. Coverage is complete, despite the limited length, and the summaries are complemented by numerous references to the original studies than can offer deeper entry into the literature.Williams then turns his attention to a brief summary of neuroscience, necessary preparation for the discussions in later chapters. This begins with descriptions of brain structure and of the methods used for assessing brain activity. The summaries of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are brief, but they do provide sufficient background for the later discussion of brain correlates of psi.

Williams raises an interesting point in this material, that the criticisms of neuroscientists skeptical of psi phenomena are commonly based on the assumption that ESP and PK must be based on some form of electromagnetic signal that carries the information or action from point A to point B. Williams notes that this assumption is considered antiquated by most parapsychologists, based on the empirical evidence that ESP and PK do not act as if they were transmissions of energy subject to physical laws. However, this weakness in the criticisms of psi, though it shows how ignorance can fuel skepticism, leaves questions of how psi actually works, and the role of brain activity, unanswered.

The next eight chapters provide the substance of the monograph, the review of research linking psi to brain activity. The methods used to assess brain activity are EEG, which measures the electrical activity of the brain from electrodes placed on the scalp, and fMRI, which provides three-dimensional images of the

activity throughout the brain. Williams organizes this material by the fundamental questions that have been investigated in research on brain correlates of psi. Is ESP associated with specific brain states indicated by EEG activity? Are the EEGs of paired individuals correlated with each other under certain circumstances? Is precognition, which involves time-shifts into the future, associated with EEG activity or the activation of specific brain regions indicated by fMRI imaging? Is ESP associated with known differences between the left and right hemispheres of the brain? Are the brains of psychic adepts different from those of other people? Is PK correlated with brain activity, based on EEG activity and on possible neuropsychological abnormalities?

One especially interesting chapter focuses on memory and the temporal lobe of the brain, suggesting that ESP is less like ordinary sensory perception and more like the recall of memories. Williams notes how telepathy, clairvoyance, and precognition seldom reveal distinct and accurate images of the target, like visual perception would, but more often yield relevant images that are in some way associated with the target, perhaps drawn from the memory store of the perceiver. This distinction, and the research supporting it, offers a novel alternative view of ESP processes that could lead to new domains of exploration.

Another chapter provides detailed criticism of a widely circulated study by neuroscientific investigators that claimed to disprove the existence of psi. Surprisingly, the claims were made based on the failure to find either successful ESP performance or any associated changes in regional brain activation in the study's participants. Obviously, changes in brain activity associated with ESP should not be expected when no ESP is demonstrated. Williams details the flaws in the study, making it clear that even sophisticated technologies such as brain imaging can be misused when the investigators are ignorant of the phenomena they study.

Williams' conclusions in the final chapter are appropriately tentative, given the current state of the literature on brain correlates of psi. He notes three limitations in this literature. Variations in experimental and statistical methods among the studies in each area make it difficult to compare the results across studies and reach general conclusions. Studies to date have also differed in their participant populations. Most have used ordinary volunteers from the general public who may never have experienced psi. Only a few studies have been made of recognized experts in psi who have frequent experiences and well-developed abilities. Williams favors the latter and suggests that studies of psychic experts will be the most productive for the discovery of brain correlates of psi. This is a reasonable point, given that psi activity must be present to observe correlated brain activity. The third limitation that Williams notes is that many of the tests have only a small number of trials, which can lead to false negative conclusions due to lack of power to detect the small effects often observed in parapsychology. But, he admits, small studies can also inflate positive psi effects and lead to results that cannot be replicated reliably. Overall, he suggests that the decades of research have produced many promising findings linking psi activity with brain function. But, he argues that much more research is needed before conclusions about the brain functions associated with psi can be drawn.

One issue that I think does not receive enough attention here is the possibility that brain activity supports some, but not all, of the activities associated with ESP or PK. As a simple example, an ESP experience, whether telepathic, clairvoyant, or precognitive, may require brain activity, or it could take place in consciousness outside of the brain. Either way, reporting the experience will almost certainly require brain-based language functions. Similarly, any memory of the experience will likely require the same brain-based activities associated with ordinary memories. Given that most experimental studies rely on memory and reporting of psi experiences, turning them into behaviors that can be recorded, assessments of brain activity no matter how sophisticated may illuminate only a small and perhaps insignificant component of the overall process. Conclusions drawn from such limited information may mislead us as to the true nature of psi and never answer the most important questions.

In conclusion, I strongly recommend this monograph to those who desire an introduction into what is currently known about the links between psi phenomena and the brain. These may be individuals who are just beginning to explore in the field of parapsychology as well as experienced parapsychologists considering the use of neuroscientific methods to further their investigations. Williams has provided a concise summary suitable for both. He is commended for his thoughtful and even-handed presentation of the

current state of knowledge, both the accomplishments and the failures, with appropriate recognition of the limitations of each. We can hope that his optimism for the future of this field will be confirmed by the work of those who are influenced by his writing.

James D. Lane

Department of Psychiatry and Behavioral Sciences
Duke University School of Medicine
Durham, NC, 27710, USA
jdlane@duke.edu

GLOSSARY

The definitions of most of the following terms have been borrowed or adapted from A Glossary of Terms Used in Parapsychology *by Michael A. Thalbourne (republished by Puente Publications, Charlottesville, VA, USA, 2003). We highly recommend this book to those who seek a more complete glossary of parapsychological terms.*

AGENT: In a test of GESP, the individual who looks at the information constituting the target and who is said to "send" or "transmit" that information to a percipient; in a test of telepathy and in cases of spontaneous ESP, the individual about whose mental states information is acquired by a percipient. The term is sometimes used to refer to the subject in a test of PK.

ANOMALOUS COGNITION (AC): A form of information transfer in which all known sensory stimuli are absent; that is, some individuals are able to gain access to information by an as yet unknown process; also known as remote viewing (RV) and clairvoyance.

ANOMALOUS PERTURBATION (AP): A form of interaction with matter in which all known physical mechanisms are absent; that is, some individuals are able to influence matter by an as yet unknown process; also known as psychokinesis (PK).

CALL: (As noun), the overt response made by the percipient in guessing the target in a test of ESP; (as verb), to make a response.

CLAIRVOYANCE: Paranormal acquisition of information about an object or contemporary physical event; in contrast to telepathy, the information is assumed to derive directly from an external physical source and not from the mind of another person.

CLOSED DECK: A procedure for generating the target order for each run, not by independent random selection of successive targets, but by randomization of a fixed set of targets (e.g., a deck of 25 ESP cards containing exactly five of each of the standard symbols).

CONFIDENCE CALL: A response the subject feels relatively certain is correct and indicates so before it is compared with its target.

CRITICAL RATIO (CR): A mathematical quantity used to decide whether the size of the observed deviation from chance in a psi test is significantly greater than the expected degree of random fluctuation about the average; it is obtained by dividing the observed deviation by the standard deviation; also called the z statistic.

Critical Ratio of Difference (CR_d): A critical ratio used to decide whether the numbers of hits obtained under two conditions (or by two groups of subjects) differ significantly from each other; it is obtained by dividing the difference between the two total-hits scores by the standard deviation of the difference.

DECLINE EFFECT: The tendency for high scores in a test of psi to decrease, either within a run, within a session, or over a longer period of time; may also be used in reference to the waning and disappearance of psi talent.

DIFFERENTIAL EFFECT: In an experiment where the subjects are tested under two different procedural conditions: (i) the tendency of subjects who score above chance in one condition to score below chance in the other, and vice versa; (ii) the tendency of one condition to elicit psi-hitting from the group of subjects as a whole and the other condition to elicit psi-missing.

DISPLACEMENT: A form of ESP shown by a percipient who consistently obtains information about a target that is one or more removed, spatially or temporally, from the actual target designated for that trial.

Backward Displacement: Displacement in which the target extrasensorially cognized precedes the intended target by one, two, or more steps (designated as –1, –2, etc.).

Forward Displacement: Displacement in which the target actually responded to occurs later than the intended target by one, two, or more steps (designated as +1, +2, etc.).

ESP CARDS: Special cards, introduced by J. B. Rhine, for use in tests of ESP; a standard pack contains 25 cards, each portraying one of five symbols, viz., circle, cross, square, star, and waves.

EXPERIMENTER EFFECT: An experimental outcome that results, not from manipulation of the variable of interest itself, but from some aspect of the experimenter's behavior, such as unconscious communication to the subjects, or possibly even a psi-mediated effect working in accord with the experimenter's desire or motivation.

EXTRASENSORY PERCEPTION (ESP): Paranormal cognition; the acquisition of information about an external event, object, or influence (mental or physical; past, present, or future) in some way other than through any of the known sensory channels.

FORCED-CHOICE TEST: Any test of ESP in which the percipient is required to make a response that is limited to a range of possibilities known in advance.

FREE-RESPONSE TEST: Any test of ESP in which the range of possible targets is relatively unlimited and is unknown to the percipient, thus permitting a free response to whatever impressions come to mind.

GANZFELD: Term for a special type of environment (or the technique for producing it) consisting of homogeneous, unpatterned sensory stimulation; an audiovisual ganzfeld may be accomplished by placing halved ping-pong balls over each eye of the subject, with diffused light (frequently red in hue) projected onto them from an external source, together with the playing of unstructured sounds (such as "pink noise") into the ears.

GENERAL EXTRASENSORY PERCEPTION (GESP): A noncommital technical term used to refer to instances of ESP in which the information paranormally acquired may have derived either from another person's mind (i.e., as telepathy), or from a physical event or state of affairs (i.e., as clairvoyance), or even from both sources.

GOAL-ORIENTED: Term for the hypothesis that psi accomplishes a subject's or experimenter's objective as economically as possible, irrespective of the complexity of the physical system involved.

MACRO-PK: Any psychokinetic effect that does not require statistical analysis for its demonstration; sometimes used to refer to PK that has as its target a system larger than quantum mechanical processes, including microorganisms, dice, as well as larger objects.

MAJORITY-VOTE TECHNIQUE (MV): The so-called repeated or multipleguessing technique of testing for ESP. The symbol most frequently called by a subject (or a group of subjects) for a given target is used as the "majority-vote" response to that target on the theory that such a response is more likely to be correct than one obtained from a single call.

MEAN CHANCE EXPECTATION (MCE): The average (or "mean") number of hits, or the most likely score to be expected in a test of psi on the null hypothesis that nothing apart from chance is involved in the production of the score.

MICRO-PK: Any psychokinetic effect that requires statistical analysis for its demonstration. Sometimes used to refer to PK that has as its target a quantum mechanical system.

NEAR-DEATH EXPERIENCE (NDE): A predominantly visual experience undergone by persons who either seem to be at the point of death but then recover, or who narrowly escape death (as in a motor car accident) without being seriously injured. NDEs often incorporate out-of-body experiences.

OPEN DECK: A procedure for generating a target order in which each successive target is chosen at random independently of all the others; thus, for example, in the case of a standard deck of ESP cards whose target order is "open deck," each type of symbol is not necessarily represented an equal number of times.

OUT-OF-THE-BODY EXPERIENCE (OBE): An experience, either spontaneous or induced, in which one's center of consciousness seems to be in a spatial location outside of one's physical body.

PARANORMAL: Term for any phenomenon that in one or more respects exceeds the limits of what is deemed physically possible according to current scientific assumptions.

PARAPSYCHOLOGY: The scientific study of certain paranormal or ostensibly paranormal phenomena, in particular, ESP and PK.

PERCIPIENT: The individual who experiences or "receives" an extrasensory influence or impression; also, one who is tested for ESP ability.

POLTERGEIST: A disturbance characterized by physical effects of ostensibly paranormal origin, suggesting mischievous or destructive intent. These phenomena include such events as the unexplained movement or breakage of objects, loud raps, electrical disturbances, and the lighting of fires.

POSITION EFFECT (PE): The tendency of scores in a test of psi to vary systematically according to the location of the trial on the record sheet.

PRECOGNITION: A form of ESP involving awareness of some future event that cannot be deduced from normally known data in the present.

PROCESS-ORIENTED: Term for research whose main objective is to determine how the occurrence of psi is related to other factors and variables.

PROOF-ORIENTED: Term for research whose main objective is to gain evidence for the existence of psi.

PSI: A general term used either as a noun or adjective to identify ESP or PK.

PSI-HITTING: The use of psi in such a way that the target at which the subject is aiming is "hit" (correctly responded to in a test of ESP, or influenced in a test of PK) more frequently than would be expected if only chance were operating.

PSI-MISSING: The use of psi in such a way that the target at which the subject is aiming is "missed" (responded to incorrectly in a test of ESP, or influenced in a direction contrary to aim in a test of PK) more frequently than would be expected if only chance were operating.

PSYCHOKINESIS (PK): Paranormal action; the influence of mind on a physical system that cannot be entirely accounted for by the mediation of any known physical energy.

RANDOM EVENT GENERATOR (REG): An apparatus (typically electronic) incorporating an element capable of generating a random sequence of outputs; used in automated tests of psi for generating target sequences; in tests of PK, it may itself be the target system that the subject is required to influence; also called a random number generator (RNG).

RECURRENT SPONTANEOUS PSYCHOKINESIS (RSPK): Expression for paranormal physical effects that occur repeatedly over a period of time; used especially as a technical term for poltergeist disturbances.

REMOTE VIEWING: A term for ESP used especially in the context of an experimental design wherein a percipient attempts to describe the surroundings of a geographically distant agent.

RESPONSE BIAS: The tendency to respond or behave in predictable, nonrandom ways.

RETROACTIVE PK: PK producing an effect backward in time; to say that event A was caused by retroactive PK is to say that A would not have happened in the way that it did had it not been for a later PK effort exerted so as to influence it; sometimes abbreviated as *retroPK*; also referred to as *backward PK* or *time-displaced PK*.

RUN: A fixed group of successive trials in a test of psi.

SHEEP-GOAT EFFECT (SGE): The relationship between one's acceptance of the possibility of ESP's occurrence under the given experimental conditions and the level of scoring actually achieved on that ESP test; specifically, the tendency for those who do not reject this possibility ("sheep") to score above chance and those who do reject it ("goats") to score below chance.

SPONTANEOUS CASE: Any psychic occurrence that takes place naturally, and is often unanticipated—psi in a real-life situation, as opposed to the experimentally-elicited psi phenomena of the laboratory.

STACKING EFFECT: A spuriously high (or low) score in a test of ESP when two or more percipients make guesses in relation to the same sequence of targets; it is due to a fortuitous relationship occurring between the guessing biases of the percipients and the peculiarities of the target sequence.

TARGET: In a test of ESP, the object or event that the percipient attempts to identify through information paranormally acquired; in a test of PK, the physical system, or a prescribed outcome thereof, that the subject attempts to influence or bring about.

TELEPATHY: The paranormal acquisition of information about the thoughts, feelings, or activity of another conscious being.

TRIAL: An experimentally defined smallest unit of measurement in a test of psi: in a test of ESP, it is usually associated with the attempt to gain information paranormally about a single target; in a test of PK, it is usually defined in terms of the single events to be influenced.

VARIANCE: A statistic for the degree to which a group of scores are scattered or dispersed around their average; formally, it is the average of the squared deviations from the mean; in parapsychology, the term is often used somewhat idiosyncratically to refer to the variance around the theoretical mean of a group of scores (e.g., MCE) rather than around the actual, obtained mean.

Run-Score Variance: The variance around the mean of the scores obtained on individual runs.

Subject Variance: The variance around the mean of a subject's total score.

INSTRUCTIONS FOR AUTHORS

Purview of the Journal

The *Journal of Parapsychology* (founded in 1937 at Duke University, and indexed in *PsycInfo* and *Scopus*) invites submissions relevant to parapsychology and related areas, including anomalous experiences, alterations of consciousness, and the nature of consciousness in general. All relevant disciplines, including psychology, physics, and biology, but also history, anthropology and other social sciences and the humanities are within the purview of the journal. Quality scholarly contributions, whether supporting the psi hypothesis or not, are welcome.

Types of Manuscripts

The *Journal* will publish these kinds of papers:

a) Empirical papers, presenting new data and/or analyses, and using quantitative, qualitative, or mixed methods. ”Empirical” should be interpreted broadly to refer to systematic inquiry including experimental manipulation of variables but also, single case designs, analysis of collection of cases, phenomenological studies, historical analyses, systematic case studies, field research, and so on.

b) Theoretical papers, discussing new theories or novel implications of theories presented previously.

c) Methodological and analytical papers, offering new or alternative ways of conducting research and/or analyzing data.

d) Debates and letters about current issues in the field and papers recently published in the JP.

e) Book reviews from the relevant international literature, which will be assigned (the JP will not consider unrequested reviews). If you are the author or publisher of a new book in the field contact the editor for more information about potentially submitting it for a review.

Submission of Manuscripts

Manuscripts must be submitted to the *Journal* electronically (as an e-mail attachment) in Word or compatible word processing program to: etzel.cardena@psy.lu.se. All submissions must be single column, double-spaced, in Times New Roman size 12. Preparation of the manuscript in all respects must follow the guidelines presented in the 6th edition of the *Publication Manual of the American Psychological Association* (APA style; of which there are many formats in the Web) and must be written at a level commensurate with academic publication. If the writing of the authors is not at this level, they might consider using a professional academic editor service, such as John Palmer’s, the former *JP* Editor at john@rhine.org, or some other *before* submitting their publication. British and American English spelling are both acceptable as long as they are used consistently.

Articles must include an abstract no longer than 200 words in a single paragraph, as well as up to 6 keywords. Do not include references in the abstract. In the case of research, the methods section must include information about the experimenters interacting with the participants (e.g., age, gender, style of interaction such as friendly, neutral, or peremptory), as well as information for all experimenters and authors regarding their level of support for the psi hypothesis using the following scale (5 -strongly supportive-, 4 -supportive-, 3 -neutral- 2 -not supportive-, 1 -strongly not supportive-). The appropriate institutional review board must have previously approved all research with living human beings and this information should be mentioned in the methods section. Footnotes are discouraged; use instead, if at all necessary, endnotes. Close attention should be paid to the formatting of references and quotations (which must follow APA style). Before submission the authors must check that all items in the reference list have matching text

entries and vice versa. Quotations should be double-checked for accuracy and their page numbers cited in the text. Statistical values should also be double-checked for accuracy.

Tables and figures must have a title or caption, be numbered, and follow APA style. Figures and photos must be submitted electronically and they cannot be in color. Resolution should be a minimum of 300 dpi. Vector art (e.g., *Adobe Illustrator*, encapsulated postscript) is preferable to bitmaps.

Descriptive statistics (e.g., mean, standard deviation) must be reported in addition to inferential statistics (e.g., *t* tests), which should also include the specific *p* value and measures of effect size (authors might consider consulting the Statistical Guidelines for Empirical Studies by Tressoldi and Utts published in the *Parapsychology: A handbook for the 21st century* edited by Cardeña, Palmer, and Marcusson-Clavertz, 2015). Although not mandatory, it is strongly recommended that all research, exploratory and even more so confirmatory, be preregistered, for instance through https://koestlerunit.wordpress.com/study-registry and that data be made available to other potential researchers through a depository such as data.world. Meta-analyses are encouraged when multiple studies have used the same variables.

Manuscripts may be refereed anonymously if so requested by the authors, in which case authors should mask their identities in the manuscript. Manuscripts accepted for publication are copyedited for grammar and style. Copyedited text will be submitted to the author for any required changes and approval, and a prompt response is required. Substantive revisions are generally not allowed after the manuscript has been accepted.

Articles in the *Journal* must be original, which means that they cannot have been published previously, either in whole in part (except for isolated paragraphs), in a journal or the Web. As a condition for publication in the *Journal*, authors of articles must sign a document assigning copyright of the article to the *Journal*, and permission must be obtained from the Editor before the article can be published or reproduced elsewhere, including on a personal website.