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.

For the first time in 2013, an electronic version of the *Journal* will be 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. Single issues (current or archival) cost $35.00 each; some back issues are no longer available. 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.

Copies of this publication are available in 16-mm microfilm, 35-mm microfilm, 105-mm microfiche, article copies, and on compact disc from ProQuest, 789 E. Eisenhower Pkwy., P. O. Box 1346, Ann Arbor, MI 48106-1346. Photocopies of individual articles may be obtained from *The Genuine Article*, ISI, 65 E. Wacker Place, Suite 400, Chicago, IL 60601. Some articles are also available through Info Trac OneFile and Lexscien.

*Copyright and permission.* Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Parapsychology Press, provided that the base fee is paid directly to Copyright Clearance Center, 222 Rosewood Dr., Danvers, MA 01923. For those organizations that have been granted a photocopy license by CCC, a separate system of payment has been arranged.

The *Journal* is an affiliated publication of the Parapsychological Association.  
Copyright © 2013 by the Parapsychology Press  
ISSN 0022-3387
Volume 77 / Number 1     Spring 2013

EDITORIAL

JP Publication Policy: Statistical Issues  
John Palmer  

PARAPSYCHOLOGICAL ASSOCIATION

2012 Presidential Address: What Have We Learned About Psi?  
Reflections on the Present of Parapsychology  
Alejandro Parra  

ARTICLES

Can Parapsychology Move Beyond the Controversies of Retrospective Meta-Analyses?  
J. E. Kennedy  

Parapsychological Experience as Anomalous Experience Plus Paranormal Attribution: A Questionnaire Based on a New Approach to Measurement  
Harvey J. Irwin, Neil Dagnall, and Kenneth Drinkwater  

A Community Survey of Anomalous Experiences: Correlational Analyses of Evolutionary Hypotheses  
James McClenon  

Performance in Group Telepathy Experiments as a Function of Target Picture Characteristics  
Jan Dalkvist  

An Anomaly of an Anomaly: Investigating the Cortical Electrophysiology of Remote Staring Detection  
Ian S. Baker and Paul Stevens  

Anomalous Remote Diagnosis: Mental and Motor Psi Impressions Under Iconic Representation of the Person-Target  
Alejandro Parra and Juan Carlos Argibay
OBITUARY

Ingo Douglas Swann
by Edwin C. May

BOOK REVIEWS

Exploring Frontiers of the Mind-Brain Relationship
Edited by Alexander Moreira-Almeida and Franklin Santano Santos
Vernon M. Neppe

Exceptional Experience and Health: Essays on Mind, Body and Human Potential
Edited by Christine Simmonds-Moore
Athena A. Drewes

Perspectives of Clinical Psychology: An Introductory Reader
Edited by Wim H. Kramer, Eberhard Bauer, and Gerd H. Hövelmann
Yolaine M. Stout

Correspondence (Neppe and Close, Potts)

Glossary

Instructions for Authors

We would like to thank the following persons for their work in translating abstracts for this issue of the Journal: Renaud Evrard (French), Etzel Cardena, (Spanish), and Everhard Bauer (German).
EDITORIAL

JP Publication Policy: Statistical Issues

Before I proceed to the topic expressed in the title, I want to officially inform readers of two broader changes in the JP. First, our electronic edition, in the past available only to members of the Parapsychological Association (PA) through the PA website, will now be available to all subscribers through the Rhine Research Center’s website (www.rhine.org), along with electronic copies of back issues. The second change, already obvious to those of you who are reading our print edition, is an increase in the page size from 6 x 9 to 8.5 x 11 inches. This is the first time the JP has ever appeared in this larger form. There are two reasons for the change. First, we save substantially on our printing costs. Second, it is easier for my Managing Editor, Dave Roberts, who also is in charge of producing our finished product, to handle the large tables and figures authors frequently submit.

On to Statistics

The PA recently adopted some radical new guidelines for the presentation of the results of statistical analyses in research papers submitted for its annual convention. Were it not for a technicality, adherence to these guidelines would have been required for papers submitted for the 2013 convention. I do not entirely agree with these guidelines, which seem to be based on trends that are gaining currency in mainstream psychology (Cumming, 2012). Nonetheless, they led me to conclude that the JP also should have statistical guidelines. I decided to present these guidelines in this Editorial, in addition to separate guidelines for authors, for two reasons. First, what I have to say should be of interest to all readers who would like to know how we determine statistically whether a psi effect is real. Second, I will be presenting my personal opinions in addition to the publication requirements; they are not always the same. Of course, it would be practically impossible to discuss every statistical issue an author might confront, so I will restrict myself to the issues I consider most important and controversial in contemporary parapsychology.

Effect Sizes

Effect sizes (ESs) are standardized estimates of the magnitude of an effect. Much has been made recently about the desirability of including them in experimental reports, even to the point of suggesting that they are a better measure of a study’s success than the traditional p-value. Essentially the same viewpoint has been espoused by parapsychology’s most prominent statistician, Jessica Utts (1988). In my opinion, effect size is the better measure of success only if everyone agrees that an effect is genuine and the real question is whether it is big enough to be of practical value. This circumstance often arises in medicine where, for example, you are concerned with the practical effectiveness of a new drug. If, for example, the drug is shown to be effective in more patients than placebo to a statistically significant degree because of a large sample size, there’s nothing to get excited about if the success rate of the drug is 5% and the success rate of the placebo is 3%. Here, size clearly matters. (However, in this and similar examples, the raw percentage would seem to be more informative than the ES; the ES seems preferable only if there is need for a standardized measure, as is often the case in meta-analysis). However, in parapsychology, the question is almost always whether the effect is real, not how big it is if it is real. To answer the “reality” question, it is the p value that should be consulted. There are several ways to legitimately describe a p value. The most relevant description to parapsychology I have found appears in a professionally written essay in Wikipedia (2013): “[a p value] is a measure of how likely the data is to have occurred by chance, assuming the null hypothesis is true.” Control issues notwithstanding, the (non)likelihood of chance is the basis on which we declare that a psi effect is real, for valid reasons that should be obvious.

Although I question the usefulness of ESs in most parapsychology papers, authors are free to use them and the corresponding confidence intervals (CIs), provided the p value is reported as well and the effect is statistically significant. Why the latter exception? When you say an effect is not statistically significant, you are in effect claiming that the ES is zero in the population. To follow this with a nonzero ES
contradicts the implication of the $p$ value. Although one might reply that the citation is justified to honor the possibility of a Type I error, the combination is close enough to self-contradictory to be objectionable. However, as stipulated in the current Instructions for Authors, authors must continue to report other descriptive statistics, such as means and standard deviations, regardless of statistical significance. Finally, and most importantly, I will never allow an ES or other magnitude estimate to serve as the sole justification for a claim that a psi effect is real.

**The Multiple Analysis Problem**

By far the most common statistical criticism I get from my referees concerns “multiple analyses.” The point is that when you perform a large number of analyses, you expect some of these to be significant by chance. Therefore, the criterion for significance for any one of these multiple effects must be made more conservative to correct the problem.

To begin with, my publication policy from now on will be that any effect that is statistically significant by the traditional criterion must be claimed as tentative unless the researcher does something acceptable to justify a stronger claim. I refer to such statements as “disclaimers.” They can apply to individual effects, but in most cases a single blanket statement can be made covering most or all of the effects reported for the experiment.

**Adjusting the alpha level.** So what can be done to avoid the disclaimer? The most common solution to the multiple analysis problem is to make the criterion for statistical significance more rigorous by applying a multiple analysis correction. The most common of these corrections in parapsychology is the Bonferroni adjustment. You simply take the original significance criterion, usually .05 in psychology and parapsychology, and divide it by the number of analyses you need to adjust for (which I call the “base $N$”), and this becomes the new criterion (alpha level) that the effect must meet. For example, if there are 10 base $N$ analyses, you divide .05 by 10 and get .005. Any effect requiring the adjustment must be $p < .005$ to be considered significant.

I have serious concerns about these multiple analysis corrections. First, there seems to be no consensus as to whether they should be applied only to unplanned post hoc effects (data snooping), or also planned post hoc effects, or even hypothesized effects. Similarly, there is no sound justification for any particular base-$N$ criterion; it’s rather arbitrary. It used to be that all you were supposed to correct for were paired comparisons inside an ANOVA table using a Scheffé test or such. Then it came along that you had to correct for many more of the analyses in the study, but there seems to be disagreement on how many more. For example, can you exclude hypothesized effects or even planned post hoc effects? Most post hoc analyses in my own research are unplanned attempts to understand some other significant effect in the study. I would argue that these secondary analyses shouldn’t count in the base $N$ because they are yoked to a primary effect. However, I’m sure this could be argued either way, and I doubt a multiple analysis correction aficionado would have much sympathy for my plea. So let’s agree to elevate the base $N$ to include all the analyses in the study. But why stop there? It seems to me that if you follow this multiple analysis reasoning to its logical conclusion, the proper base $N$ (at a minimum) should be all the analyses ever conducted in the social/behavioral/neurophysiological sciences, which of course would guarantee that no analysis could ever be claimed as significant.

However, my main objection to multiple analysis corrections is that the whole notion that the de facto objective likelihood of an effect being real is influenced by how many other analyses an investigator decides to conduct is absurd on its face (unless, of course, one assumes a paranormal cause!) Demanding the correction could also discourage investigators from conducting exploratory analyses as a way to protect the significance of the more important analyses. This is especially a problem in parapsychology because the great majority of our studies are primarily exploratory, and we need good hypothesis generation. They also markedly increase the Type II error rate. Finally, I have a particular problem with the Bonferroni because it assumes that the analyses are independent; this assumption usually is grossly violated in practice.

So, my policy on multiple analysis corrections is the following. Authors are free (but not required) to apply them to however many analyses they wish, so long as they specify the base $N(s)$ and describe any classes of analyses that are excluded. However, these corrections cannot be used to avoid the disclaimer.
Replication. So, what’s the alternative? My position is that to avoid the disclaimer, any initially discovered significant effect must be successfully replicated, ideally (but not necessarily) by an independent investigator. An important virtue that replication supplies, which a huge $p$ value for the original effect cannot supply, is evidence of generality or robustness, that the effect is not beholden to a particular sample or set of experimental circumstances. This is particularly true in parapsychology, where the effects are subtle or elusive, and we don’t have a very good idea of all the factors that can influence an experimental outcome. Even a replication by the same investigator is different than the original study, if for no other reason than the experimenter will approach it with a different attitude, caused by the success of the original study. Moreover, the sample of participants can be unexpectedly different in some crucial respect (see, e.g., Stanford & Frank, 1991). Because of this principle of generality, I would be more impressed by the combined statistical significance of 10 distinct parapsychology studies with 50 participants each than a significant result from a single study with 500 participants; multiple studies are better.

So how do we define a successful replication? My preference is a traditional criterion, statistical significance at the .05 level, one-tailed. However, I sense that there is a lack of agreement in our field about this criterion, so I will allow a more liberal criterion in the JP. At this time, the most liberal alternative I am willing to accept is combined significance of the original and replicating study using the Stouffer $Z$, as illustrated in the study by Dalkvist in this issue. What I will not allow is successful replication to be claimed if the effect size (ES) of the replication study falls within the 95% CI of the effect size of the original study. Although application of this criterion might be defensible in other fields, where one is seeking a reliable point estimate for the magnitude of an effect, in our field it has the disastrous consequence of allowing an investigator to claim evidence for psi regardless of the outcome of the analysis: If the Study 2 ES falls within the Study 1 CI, the investigator claims a successful replication; if it falls outside this CI, the investigator claims a statistically significant difference between the two studies. Imagine the fallout if our critics ever caught on to this!

In my opinion, replication should be required for any effect regardless of (a) whether it was unplanned post hoc, planned post hoc, or hypothesized, (b) the $p$ value (adjusted or unadjusted), and (c) the sample size. I think I’ve explained my reasons for (b) and (c) adequately above, but (a) needs more comment, particularly with regard to hypothesized effects. The claim that hypothesized effects should be treated more leniently than other effects is the flip side of the statement routinely uttered by critics of parapsychology that “extraordinary claims require extraordinary proof.” Many years ago, I wrote a paper attacking this claim, arguing that the criteria an effect must meet to be accepted as a scientific fact should be uniform, at least within a broad scientific domain such as psychology (including parapsychology; Palmer, 1987). To give preference to findings because they are consistent with some hypothesis or theory tends over time to unfairly bias the literature in favor of that hypothesis or theory. In the present case, this problem would manifest by declaring as real an unreplicated effect hypothesized by, and thus consistent with, some theory, while labeling the same finding, if it did not support (and indeed might be inconsistent with) the theory, as only tentative.

However, for JP publication policy, I will again at least partly bow to what I perceive as the more commonly held view within parapsychology and adopt a criterion more liberal than my personal preference. Thus, I will allow significant hypothesized effects, but not significant post hoc effects (planned or unplanned), to escape the disclaimer, which should read something like “The finding(s) is (are) tentative pending replication.”

Registration of Experiments

In a recent letter to the JP, Caroline Watt (2012) announced that the Koestler Parapsychology Unit of the University of Edinburgh was setting up a registry to which researchers are invited to submit their hypothesized and planned analyses in advance of data collection. The purpose of the registration is to provide concrete evidence that these analyses are “planned prior to conducting the experiment” (p. 403). Insofar as authors abide by my preference that nonconfirmatory hypothesized and planned analyses be treated the same way statistically as unplanned analyses, they need not submit such analyses to the registry. On the other hand, I think the registry is a good idea for other analyses. My only caveat is that I think the
registry would have more credibility among researchers and achieve maximum usage if it were sponsored by the PA rather than any individual laboratory. I would only consider registration by the previously defined “unencumbered” researchers to be a requirement for publication in the JP if the same policy were agreed to by the editors of all the major journals in the field.

Power Analyses

Statistical power is defined as the (a priori) probability that a false null hypothesis will be rejected. Its primary value is to alert investigators how large their sample must be if they hope to have a decent chance to successfully replicate a previously obtained effect. This is a sobering exercise, because researchers are likely to discover that the needed N is larger than they intuitively expect and can readily obtain in practice. My policy for the JP will be to recommend, but not require, that the power statistic be reported in cases of replication. However, I consider power analysis potentially problematic for effects other than replications. The reason is the need to insert an ES estimate in the power analysis formula. In the case of a replication, the solution is straightforward; you insert the ES from the original study. However, for other effects, particularly exploratory ones, there is often no sound basis for estimating an ES. I ran into this problem myself a few years ago when I submitted a psychology article to a mainstream journal. The editor originally insisted that I publish a power statistic, even though the effect was unprecedented in the literature and I didn’t have a clue how to estimate an ES for the analysis. Fortunately for me, the editor relented and the paper was published without a power statistic (Palmer, Mohr, Krummenacher, & Brugger, 2007).

For the JP, if authors want to report a power statistic in nonreplication cases, they must provide a sound justification for the ES they insert in the power analysis formula. Moreover, authors must use extreme caution in attributing the failure of an effect to reach significance to low power, even if the ES is large. With low Ns, effect sizes can be highly unstable, and a high one could easily be the luck of the draw. The burden of proof falls on the researcher to conduct a large enough experiment to obtain a significant p value.

Conclusion

One of my perks for being a journal editor is that I get to write editorials in which I can say pretty much anything I want without having to worry about my remarks being vetoed by some referee. The current editorial is of course a case in point. However, readers are free to express their disagreement with my views in a Letter to the Editor, which I will publish so long as it’s in good taste. Even if there are no Letters, I hope my remarks will provoke thought and discussion about these matters among the readership. That’s always a healthy development.

References


John Palmer
WHAT HAVE WE LEARNED ABOUT PSI?
REFLECTIONS ON THE PRESENT OF PARAPSYCHOLOGY

By Alejandro Parra

Saint Augustine, the celebrated theologian of the Catholic Church, tells us that, as he was walking on the shore of a beach meditating on the mystery of the Holy Trinity, he saw a child playing with a nutshell. Saint Augustine got close: “What are you doing, little one?” he asked the child. “I am trying to collect all the water of the ocean in this nutshell,” he answered. The theologian, looking at him with surprise but with compassion at the same time, asked, “But how could you pretend to collect all the water of this immense ocean in this small nut?” With plain candor, the child responded, “And how do you pretend to understand the marvelous mystery of God with your tiny head?” (Brown, 2000, p. 213). It’s the same as it was for Saint Augustine: For those of us who work in parapsychology, psi escapes our comprehension and control. We parapsychologists confront one of the greatest problems of human nature: Is psi an ontological dimension (related to being), and as a consequence, still basically unexplored?

In my judgment, one of the great merits and achievements of modern parapsychology is to have applied the rigor of the scientific method to the exploration of phenomena which, even today, continue to attract great popular interest. These subjects always generate much metaphysical speculation. However, their study also comes accompanied with feelings of rejection, scorn, or direct negation on the part of the rest of the scientific community.

The systematic study of spontaneous cases—such as reports of psi experiences and poltergeists; the problem of the survival of the human personality after the death of the physical body in the form of memories of past lives by children; spirit identification employing sophisticated codification of messages; apparitional encounters, the investigation of anomalous experiences which for centuries have traditionally been the patrimony of occultism, spiritism, and folklore (“astral trips,” luck, detection at a distance, healing by faith, psychic reading); and the therapeutic treatment of traumatic anomalous experiences—all show the creativity in the designs of the parapsychologists, often surpassing what one encounters in other social sciences. Of course, many of these experiences have also been explored in attempts to identify the positive and significant correlations of psi with psychological, physiological, physical, and even geomagnetic variables.

But this conquest has also had a high cost in terms of the comprehension of these phenomena and experiences. The scientific method guarantees us security, control, and dominion, but at the same time it limits us, it narrows us, and it subjects us to a partial vision of the nature of psi phenomena. Its negative counterpart, a suffocating obsession for the scientific method, may work against psi performance if we were to take seriously into account all the recommendations of the skeptics to accommodate ourselves to their (pre)judgments. Social scientists with weak and immature scientific training can easily become prey to this “irrational” rationalism and mitigate their interest in the exploration of these incompletely understood human experiences, no matter what their ultimate nature may be. They compel us to adopt an absolutely conceited position like the one the skeptics attempt to construct for us with their arguments.

Even so, it is not as if we are trying to reject the scientific method (Feyerabend, 1988), fight the skeptics, or burn in a bonfire our parapsychology laboratories. It is notable that, in many studies carried out by the pioneers of psychical research, some of them ended up asking themselves about the nature of psychic experiences. Positivist scientists, like the French physiologist Charles Richet, were themselves convinced by their discovery of much of psychic mechanics, although the “explicative theories” proposed by these pioneers did not totally negate their previous vision of the world. For example, Richet’s theory of the “sixth sense” (1928) was supported by principles entirely physical and physiological, while he hoped that the next generations would respond to those questions that, even today, paradoxically, we continue to formulate in our laboratories.

1 This article is the Presidential Address delivered at the 55th Annual Convention of the Parapsychological Association, Durham, NC, USA, August 9–12, 2012.
At the beginning of the 21st century, many investigators discovered that highly emotional events were associated with spontaneous psi experiences. They initiated a systematic study of these relationships using many objective types—for example, drawings or other hidden objects—with the cooperation of psychics, mediums, and common persons. But these simple experiments were severely criticized for lack of controls. Consequently, many of them remained marginal in the modern parapsychological literature. Even when they had successful results, they demonstrated aspects that many of us would have liked to replicate in our own laboratories.

The historical and social context has much to say in this respect. Following the two devastating world wars, much of the good psychical research, mainly in Europe, was silenced, predominantly replaced up to the present by research based on the Rhinean model. I am under the impression that this change of course led to the emergent Rhinean model becoming the only course. Attention to the quality and quantity of these investigations, which could have been conducive to a more sensible and relevant understanding of psi, was not sufficiently taken into consideration.

When the scientific method was applied with rigor by modern parapsychology absent a predictive theory, instead of demonstrating how much information we have managed to discover about psi, it made evident our ignorance about the phenomenon every time we had to express ourselves in terms such as “anomalous cognition” or “remote influence at a distance.” Instead, we should remember the courageous and energetic discourses published by pioneers such as Charles Richet, Oliver Lodge, and William Crookes at their academies of science, and the work of other notable scientists who defied the opposition of a great number of their contemporaries (Boirac, 1917; Crookes, 1903; Hyslop, 1905; Richet, 1923). I am sure that they would have rejected the idea of modifying the term “parapsychology,” looking for substitutes that could be better adapted to the theoretical vacuum that we are leaving in our articles, books, or texts in psychology and other social sciences.

Epistemology is a branch of philosophy that, basically, teaches us if our instruments of exploration are valid enough to justify knowledge. I am convinced that it is necessary to explore psi using instruments complementary to the scientific method. It is perfectly possible that phenomenological, transdisciplinary, psychoanalytical, semiological, linguistic, and systemic methods would provide us with valuable, reliable, and rigorous information—if they were suitably applied—just like the classical scientific method with which we have been attempting to this very moment to explore and understand psi.

Accidentology is a good example of a transdiscipline, due to the fact that its research methods are in close relationship to parapsychological procedures. Accidentology is of potential interest to physicians, nurses, engineers, architects, and insurance companies, because it studies the causes of accidents and develops procedures to prevent them, or—if they happen—to find strategies to confront them in time. This way the victims can be attended to with the greatest possible speed, avoiding deaths due to deficits of organization in situations of great physical and emotional tension, where other people are trying to help them. Naturally, nobody is “prepared” for an accident; similarly, nobody is prepared to experience a spontaneous psi occurrence.

The accidentologists have also examined the reliability of their observations and preventive measures in the form of simulations and controlled laboratory experiments. Inter- and trans-disciplinary action is necessary, combining strategies and experiences originating in fields so apparently opposite as nursing and engineering, just to mention those most strongly implicated. Police officers, firefighters, and security agents have contributed empirical data, resulting in a vast knowledge base of the causes of accidents, which today helps thousands of people in natural disasters (fires, landslides, earthquakes, hurricanes, storms), in transportation (cars, planes, trains, ships), and in job-related accidents of all kinds. They have explored accidents in their natural environments, have uncovered relevant statistics, and have meticulously examined physical, physiological, environmental, and psychological factors related to the principal causes of accidents.

These results point to conditions of natural observation that may contribute information about all the causes of the occurrence of psi phenomena. Parapsychologists have initiated a similar approach to exploring spontaneous psi experiences, determining their causes in an attempt to reproduce conditions that facilitate them in an appropriate experimental framework. But contrary to accidentology, the intensity of psi phenomena in natural conditions is dramatically diminished in the artificiality of the laboratory: The
mechanization of the objectives, the subjects’ and experimenters’ motivations, and the creativity of designs aimed at avoiding the controls against sensory clues or fraud working against psi performance—all these must occur in a harmony that is often difficult to sustain.

However, the controversy does not arise because of parapsychology’s methods, nor even because of its instruments, but because of its object. What is psi? What are its limitations? Disciplines such as psychoneuroimmunoendocrinology (PNIE) present epistemological problems. From a transdisciplinary perspective, PNIE explores the interrelationship of the defense mechanisms of an organism and the psychological characteristics that increase or decrease the possibilities of recuperation for a patient. A great number of successful studies include strategies of visualization, concentration, and relaxation for the treatment of biological dysfunctions such as cancer. In spite of their results, biologists do not agree about the object of study of PNIE. In which discipline should it be placed? A unidimensional treatment is impossible if we accept that the parts of the system complement and interact with one another. PNIE, like parapsychology, can also be seen, speaking epistemologically, as a transdiscipline that explores, at the same time, a psychological and physical dimension of psi (ESP and PK).

The phenomenological method is also incisive, but, unfortunately, seldom employed in parapsychology. Nevertheless, its existential humanistic appreciation of the human experience permits a vision more from the point of view of the subject having the experience than other psychologies of behavioristic or psychoanalytic orientation (Irwin, 1994). This method may allow us to resignify the paranormal (or anomalous) experience and to become sensitive to its subjective qualities that the scientific method has only limited options to investigate. Besides, it has the potential to become especially useful as a therapeutic tool for persons who have had traumatic anomalous experiences, following the model of Rogerian and transpersonal therapies (Gómez Montanelli & Parra, 2008; Grof & Grof, 1989; Kramer, 1993; White, 1990, 1993).

The border between a psychological experience and a paranormal experience is quite confused and even irrelevant for people emotionally moved by an event that escapes their everyday experience and their construction of reality (Milton, 1992; [L.E.] Rhine, 1981). Parapsychologists with experience in laboratory investigation are sufficiently competent to choose phenomenological variables for their investigations, but these variables are not always reported in parapsychological experiments, for various reasons. Extrasensory communication, out-of-the-body experiences, near-death experiences, apparitions, and/or spiritual contacts, all provide valuable information about the frequency, type, teleological significance, and subjective spatial and temporal perceptions that may be of use to elaborate proofs of hypotheses for future experimental designs in controlled laboratory conditions (Alvarado, 1984; Ring, 1984; Stevenson, 1995a, 1995b).

Linguistics is another seldom-approached field by parapsycholoists, but one of great importance for understanding the nature of psi and directing much of its behavior. The linguist Noam Chomsky (1972) conceives of language as an innate structure of the human species activated by an adequate stimulus from the social environment. Some linguists consider that, since birth, telepathy and language are one and the same (Guilfoyle, 1998). Consequently, in the bipolar universe of linguistics, the possibility that a child could paranormally absorb a structure so complex as a language’s grammar is unthinkable: At one pole, children obtain their knowledge of language via surrounding adults’ induction; at the other, children have an abstract knowledge of the language when they are born. However, some linguists consider that children possess telepathy since birth, because they are lacking a language that would interfere with their paranormal receptivity (Fitz, 1961; Schwarz, 1971).

Language functions as an isolating barrier against the telepathic exchange between people and as a substitute for telepathic exchange, since what separates the adult from the child in this instance is the presence of language. When we speak, listen, read, and write, we are virtually prohibited from using our telepathic potential. Consequently, it is possible that the success of psi depends on the semantic relations that are established in communication via telepathy, which permit us to understand the content of a message, always subject to inadvertent distortion by an incorrect interpretation. Such semantic connections are essential if we want to understand why emotionally charged events have a potentially greater “meaning” than abstract symbols or models with enormous semantic limitations such as those employed in parapsychological laboratories. A parapsychological exploration of the mother-child relationship and the elementary biological needs for survival (breathing, nourishment, affectivity) is strongly implicated in the
understanding of psi; perhaps in the future more introspective analyses will appear about the relationship between language with its rules and other paranormal experiences.

In the form of psi communication, this possibility can be examined more clearly. Radiesthesists usually employ instruments to facilitate the paranormal detection of lost objects. When the radiesthesist is localizing a target on the map of a territory, the “connection” between the radiesthetic rod or the pendulum and the target is not physical but semantic (linked by the symbolic map that the dowser sees; Levin, 1999). Paranormal detection using an object from the person from whom it is wished to obtain information, also called psychometry, operates under the same principle, even if the person is dead.

Closely linked to language, systemic theory also includes a model for the comprehension of psi. Systemic theoreticians propose that communication is a permanent social process that integrates multiple modes of behavior (gestures, looking, mimics, interindividual space) with a singular significance. Every phenomenon is meaningful, depending on its relationship with its context. Instead of an energetic transference, the systemicsists propose that communication consists of information transfers of different orders of complexity. According to this model, there are guides that order the interaction of each system with, in particular, its significance in the context and any dissonances and expectations. Systemic theory criticizes the behaviorist model of “stimulus-response,” arguing for a bidirectional interaction: “I have trained my experimenter,” says a rat to another in a laboratory. “Every time I press the lever, he gives me food.”

Thus, in a fruitless attempt to isolate the variables associated with psi phenomena, (despite the fact that they are the traditional skeptics), many parapsychologists seem to forget the necessary ecology that must reign in the experimental situation, namely: a harmonic relationship among target, participants, and internal and external variables, where the experimenters often exclude themselves from the object of their study, observing from outside the nature of a phenomenon that, in reality (and I think, intuitively) incorporates them into the scenario more as protagonists than as spectators in the theater of the mind.

As a consequence, from the systemic perspective, parapsychologists should pay more attention to the relation between repeatability and the “experimenter effect.” For example, in general terms, the lack or difficulty of repeatability in parapsychological experiments prevents experimenters from arriving at substantive conclusions. The psychology of the experimenters in the experiments is intimately linked with their negative or positive results and with their hypotheses, even when all the participants receive the same motivation from them. Future designs could experimentally examine this variable—that is, by employing a personality questionnaire unknown to the experimenters, which would permit comparing the results of “successful” psi experimenters with those deemed “not successful” (Parker, 1977; Sargent, 1980), or comparing the results of experimenters who design and administer their investigations with those of experimenters whose investigations are designed by them but administered by others, indifferent to or little motivated by the experimenters’ hypotheses.

In any case, an underlying problem remains: The lack of a general theory of psi may lead parapsychologists to elaborate a suffocating number of hypotheses, or still worse, to find foolish correlations simply because we ourselves have no agreement on whether to continue investigating according to those studies that promise a more holistic understanding of the psi process. Instead, in my personal experience, we listen attentively to the nurturing conventions of the Parapsychological Association, whose speakers present studies of high methodological quality, attractive design, and reliable controls but remain committed to aspects of psi that are completely unconnected or very little interconnected and lack a general theory that would support them.

Moreover, many social sciences count on an important, sure, and predictable subvention system [of financial support], a small part of which represents—perhaps—the total investment in the whole history of parapsychological investigation. Why do we feel so satisfied with these subventions as a positive and unrejectable economic resource? We should no longer expect to profit from the existential panic of some multimillionaire anxious about the destiny of his soul after death, or the inheritance of some parapsychology sympathizer, or the possibility of guessing the winning number of the lottery, with whose prize we would be in a position to support parapsychology for the rest of our lives and the lives of the next generations of parapsychologists in our institutes and research centers. Do parapsychologists suffer from the narrow view that the only sure path to success is to acquire appetizing fortunes that could be destined to respond to the millenarian questions that humanity has considered since antiquity? Maybe the
What Have We Learned About Psi? Reflections on the Present of Parapsychology

Governments and scientific foundations of all our countries should worry seriously about these questions and formally finance our investigations!

I define “parapsychology” as a discipline that principally applies the scientific method with the object of understanding psychological processes that operate in two forms of psychical interaction: one subjective, which permits people to obtain or to transmit information (extrasensory perception or anomalous cognition), and the other objective, in the form of a remote mental influence (psychokinesis). ESP as much as PK escapes—apparently—the control of our present dominant monist-materialist philosophical paradigm, if I use the words of philosopher Thomas Kuhn (1962/1996). After more than a century of psi research, parapsychologists have concluded that the psi phenomena produced in controlled environments are not sufficient or intense enough to be accepted as a fact of the natural world or to force a change in the present paradigm.

What have we parapsychologists learned? What do we know about psi? Which parameters should we consider “scientific”? Like Wilber (1990), I think that there exist at least three domains of human nature. There is a domain of the sensorial and physical that can be accessible by the *eye of the flesh*; a mental domain, whose ideas, thoughts, and images are perceived by the *eye of the mind*; and a domain of the transcendental or spiritual, known through the *eye of the spirit*. Each eye reveals a different aspect of reality; moreover, what is revealed by one eye is not necessarily accessible to the others. Wilber also reminds us that we make all kinds of errors when we attempt, imprudently, to see everything through only one eye or permit only one form of vision to usurp the domains of the other two.

“My shoe does not fit the foot of someone else,” says Wilber (1990, p. 18). However, there exists at least one point of contact among the three realities. Science—in the form of psychical research and parapsychology—can find aspects of reality that are of value to those interested in spirituality. To get close to a complete image of reality, we must have, at least, a tryptic vision. We can consider that there exist different forms of knowledge to approach the “real.” None is better than the other. Whether we speak of science, philosophy, or religion, each one interprets reality in ways that are its own, with different objectives and different themes.

The scientific method is a way to obtain and validate knowledge. Rigorously, parapsychology does not use an “alternative” methodology, since it applies the classic scientific method. Like Fisher, the prestigious American mathematician who defended Rhine, maintaining that parapsychology must be criticized from other places than statistics, we also can say that, if parapsychology can be questioned, it must be from some other place than the lack of application of the scientific method.

The principal problem of parapsychology originates, in my view, from its difficulty in formulating a predictive theory of psi. There is no agreement among parapsychologists about this. In spite of the existence of diverse explicative theories, the majority of them do not have “proof weight.” And even if, using meta-analysis, some empirical generalizations could be “proved,” these would have scant predictive power, being able to explain only a minimal portion of the variance.

To argue that meta-analysis permits an empirical generalization equivalent to a proof weight is an error, unless its predictive power were sufficiently strong and consistent to allow it to explain the variance. The philosophical paradigm prevailing in a determined historical or social epoch influences what we consider normal or “real,” or adjusts to our vision of reality and the mechanics that we know. If psi phenomena did not compromise the stability of the present paradigm, along with the feeling and thinking of the scientific community, probably there would not be any questioning.

The impossibility, at least for the moment, of a practical application of psi abilities also limits its acceptance. But the lack of repeatability is not an excuse, even if practical application were the only requisite for its acceptance. The problem of repeatability is a proper concern of the social sciences in general, many of which still see this requisite of the physical sciences and mechanics as a legitimate determinant of accepted knowledge. However, there are many scientists who resist parapsychological work, not only arguing that it condones low methodological quality but also admitting that the existence of psi conflicts with their vision of the world. Often, such an absolute opposition to the existence of parapsychological phenomena means that those parapsychological reports that authors look to be printed in the social or physical sciences journals, even absent any methodological defects, are rejected anyway if their results are positive.
Sociologist Marcello Truzzi (1987), a keen analyst of skepticism, elaborated on the intellectual narrowness demonstrated by the critics of parapsychology: “Scientists are not the paragons of rationality, objectivity, openmindedness and humility that many of them might like others to believe” (p. 3). Nobelist James D. Watson, codiscoverer of the structure of DNA, agreed: “One could not be a successful scientist without realizing that, in contrast to the popular conception supported by newspapers and mothers of scientists, a goodly number of scientists are ... narrow-minded and dull....” (Watson, 2010, p. 234).

As psychologist Hans Eysenk (1970, p.12) observed, “Scientists, especially when they leave the particular field in which they have specialised, are just as ordinary, pigheaded and unreasonable as anybody else, and their unusually high intelligence only makes their prejudices all the more dangerous....” The question is, are these reactions violations of prejudice or violations of the laws of nature? The materialistic assumptions that underlie the denunciations by parapsychology’s critics are already being abandoned. During the 20th century, the goal of neuroscience was to understand the workings of the mind in terms of the physical laws governing the material brain. It was an article of faith that a thorough understanding of the brain’s atoms and molecules would lead to an understanding of consciousness itself. As astronomer Carl Sagan said, “[The brain’s] workings—what we sometimes call mind—are a consequence of its anatomy and physiology, and nothing more.” (Sagan, 1986, p. 68). Or, as Nobelist Francis Crick observed, “... a person’s mental activities are entirely due to the behavior of nerve cells, glial cells, and the atoms, ions, and molecules that make up and influence them” (Crick, 1995, pp. 213–214).

The bankruptcy of the materialistic approach to consciousness is now being openly admitted. The theoretical biologist and complex systems researcher Stuart Kauffman put it, “Nobody has the faintest idea what consciousness is.... I don’t have any idea. Nor does anybody else, including the philosophers of mind.” Also, philosopher Jerry A. Fodor expressed a similar opinion, saying, “Nobody has the slightest idea how anything material could be conscious. Nobody even knows what it would be like to have the slightest idea about how anything material could be conscious. So much for the philosophy of consciousness” (both quoted in Dossey, Greyson, Sturrock, & Tucker, 2011, p. 132). Theoretical physicist Freeman Dyson (2011) agreed: “The origin of life is a total mystery, and so is the existence of human consciousness. We have no clear idea how the electrical discharges occurring in nerve cells in our brains are connected with our feelings and desires and actions.” In other words, parapsychology’s findings violate not the laws of nature but the ingrained prejudices of its critics about how the world should work. Brian Josephson (2008), a Nobel physicist at Cambridge University, is among the physicists who have carried out probed experiments in which physiological changes occur in the subject before the stimulus happens.

We can find contradictions between their occurrence and our culturally accepted view of reality, but not, as many of us have believed, between their occurrence and the scientific laws that have been so laboriously developed. However, parapsychology’s findings and the accumulated data from a century of consciousness research suggest that scientists’ buoyant pronouncements represent not understanding but a congealed intellectual enterprise that has foundered on its own inertia. The result is the illusion of understanding.

Materialists do sometimes change their views, at least somewhat. Daryl Bem in 1994 spoke with Cornell fellow faculty member Carl Sagan—a popularizer and communicator of the space and natural sciences, famous for his 1980 television series Cosmos: A Personal Voyage. Sagan was also a prominent cheerleader for a materialist view of consciousness and a brutal critic of parapsychology. As Bem and Sagan talked, Sagan repeated the skeptics’ perennial complaint that there are no replicable findings in parapsychology. Bem asked Sagan whether he had taken the time to look at contemporary research findings. When Sagan admitted he had not, Bem suggested he do so before continuing to make such an assertion. Sagan promised he would, asking Bem to send him a copy of a research paper Bem had recently completed with Charles Honorton, Director of the Psychophysiological Laboratory in Princeton, New Jersey (Dossey, 2011). Bem did this shortly afterward. The subject of the paper was a meta-analysis of a host of ganzfeld studies, in which an individual who was experiencing mild sensory deprivation attempted to describe information being sent to him in ways that did not involve sensory mediation.

The paper made an impact. Sagan called Bem and invited him to present the data and arguments to Sagan’s senior seminar called “Critical Thinking.” Bem complied shortly thereafter. The next thing
Bem heard of their discussion was the following passage, which appeared in Sagan’s (1995) last book, *The Demon-Haunted World*:

[T]here are three claims in the ESP field which, in my opinion, deserve serious study: (1) that by thought alone humans can (barely) affect random number generators in computers; (2) that people under mild sensory deprivation can receive thoughts or images projected at them; and (3) that young children sometimes report the details of a previous life, which upon checking turn out to be accurate and which they could not have known about in any other way than reincarnation. (p. 302)

If this sounds like a radical conversion, it was not, for Sagan’s next words were: “I pick these claims not because I think they’re likely to be valid (I don’t), but as examples of contentions that *might* be true. [These] ... three have at least some, although still dubious, experimental support. Of course, I could be wrong.” Although a minimal concession, it was at least a departure from dogmatic absolutism. It’s Sagan’s “might” that’s important.

The conflict over parapsychologists’ findings reveals an unfortunate development: Science, which fought for centuries to free itself from the dogma of the Church, is now mired in its own dogma, scientism. To my mind, parapsychologists’ experiments are the modern equivalent of Galileo’s telescope, down which the authorities refuse to peer; indeed, recall the learned men of Galileo’s time who refused to look through the telescope. They were of the opinion that data from telescopes were not relevant. The same thing is happening today, except that the limiting doctrine is coming not from the Catholic Church, but from science—the new religion of the 21st century.

In this suggested new approach, emphasis would be placed on identifying these nonevident target qualities by focusing on the nature and congruence consensus of percipient responses to various target events. Whether psi is able to access such qualities is quite testable, by the careful study of such responses and avoidance of confounding by possible stacking effect and response bias. To give a very simple example, one might study the existence and nature of the conventionally invisible human aura by having multiple viewers, tested independently, observe the space around the periphery of the body in which auras are hypothesized to exist, while the rest of the human target’s body is completely blocked from vision, using randomized sequences of target person present or absent, behind some truly opaque screen. Conventional free-response methods could be used in such studies. The additional use of the Projective Differential could allow both quantitative and qualitative assessments, along with quantified response congruence measures. Similar approaches can be employed, emphasizing responses and degrees of response agreement, to explore many of the additional psi functions suggested above.

Over the history of parapsychology, many calls have been made for the “critical” experiment that would at last allay all doubts about the reality of psi phenomena. Skeptics have persistently demanded ironclad research protocols. Meta-analysis has shown successful replication in several separate protocols with high confidence. Such academic battles and their accompanying publicity have had some value.

Parapsychology research has far better research protocols as a result, indeed, better than several mainstream sciences, and those protocols should give anyone who actually knows the literature confidence that, in Gertrude Stein’s pithy phrase, “there is some there there.” Of course skeptical vetting is critically important to good research. It is fair to say that we all have biases, and without help from our skeptical and critical friends, we make mistakes and overlook possible misperceptions and misinterpretations.

Given that there are many experiments and observations of high quality showing anomalies in a wide range of disciplines, along with independent findings pointing to effects of consciousness that are not accounted for in ordinary psychological or physical theories, we can say that there is excellent “evidence” that consciousness interacts with physical reality. There is a powerful general point to be made from the psi literature.

We have laboratory experiments on extrasensory perception, clairvoyance, psychometry, psychokinesis, and more. But with more than 100 years of research by highly qualified scientists looking from different perspectives at the extended capacities and limitations of the mind, we can consider whether their findings converge. We also have extensions of these efforts to learn something in the real world,
some pragmatic and some purely experimental. Pertinent to our theme, such work may be regarded as applications of techniques and findings from controlled laboratory research. Similarly, studies of micro-psychokinesis in the laboratory have led to field research on group consciousness, attempting to confirm that special states of resonance or coherence reportedly stimulated by ritual, music, collaboration, and cooperation may have a detectable presence beyond the experiential.

The Global Consciousness Project is a multilevel example of convergent evidence (Nelson, 2008). Its application of powerful modeling and statistical techniques to search for structure in this large and complex database seeks convergent evidence internally. But this evidence converges with and extends the field studies of group consciousness and laboratory research with individuals. The GCP results say essentially the same thing as do the results of decades of psi research in laboratories around the world: Consciousness is real.

Consciousness has a role to play as a presence in the physical world. Our work as psi researchers is to go on with efforts to learn more about that presence, and to make clear that the role of consciousness in the world is both real and important. In this second decade of the 21st century, it is becoming apparent that that role is critical.

Presentiment research requires us to set a goal. What would we consider a success for parapsychology and presentiment research, in particular? Ultimately, as with any science, the goal is to gain a better understanding of “life, the universe and everything” by testing and revising hypotheses and theories. To increase the pace at which we go forward, we need more people and more money. There are already a number of encouraging developments resulting in more scientists listening to, and becoming involved in, parapsychology.

The scientific method relies heavily on our present interpretation of time and causality. If empirical data and theoretical developments in presentiment research lead us to conclude that time and causality differ fundamentally from what we hold them to be now, then the scientific method may find itself, paradoxically, in uncharted territory. Unfortunately, much of parapsychological research is as far removed from this as it is possible to be, given our preoccupation with precluding fraud and error. With so little effort to make our tasks personally relevant to our participants, we open ourselves up to displacement effects. With this in mind, perhaps we should not be asking why our replication levels are so low, but rather whether we get more replication than we deserve.

Parapsychology’s real problem is not with our phenomena: It is that there is almost no money in the field. It is virtually impossible to prosper as a full-time research-contract worker in parapsychology, and many of our number have to take other jobs to subsist, especially in Latin American countries. The current situation is fairly typical. This impression is consistent with Sybo Schouten’s (1983) well-known calculation that the number of person-hours invested across the lifetime of parapsychology from its beginnings, with the establishment of the Society for Psychical Research in 1882, equates to only 2 months of research in conventional psychology in the United States.

This strategy of taking other jobs can create opportunities for graduates with an interest in parapsychology to get their first step on the academic ladder. But when the principal investigator has a university salary, then more of that money can be used to service direct research costs rather than paying wages. For example, previous grants from the Bial Foundation have been used with some success to bring PhD students into parapsychological positions, and some of these have now graduated and secured tenured positions at other universities, despite their research interests primarily involving parapsychology.

This pattern also occurs in other disciplines, but with their greater numbers they include many able technicians who are willing to conduct the kinds of modest replication extensions that Kuhn would have called “normal science.” It also seems likely that the limited number of persons involved in parapsychology has consequences for the nature of the research that is undertaken. For example, parapsychology seems to attract innovators who have been successful in developing new protocols, adapting methods from other areas, or demonstrating “proof of principle” by reporting significant psi effects using such methods. These innovators relatively quickly lose interest in simple confirmations and move on to develop yet more methods and approaches, with the early adopters soon following suit.

As the PA reaches the beginning of the second decade of this century, we are faced with a variety of issues related to the function, purpose, and development of our Association. In this short message I want
to emphasize co-operation and involvement, but I do not want to promote the illusory view that there are no real differences in approach and method or that those differences are not important. Our differences are fundamental to our identity as a profession and cannot be dismissed by mere calls for union that ignore the roots of our differences and offer little in the way of specific solutions to our problems. What I propose is the constructive use of our differences to serve our best interests. Our disagreements signal our strength, because they highlight the different areas of expertise that have been brought to bear on the complex problems of our field.

We can all offer much to effect future change. The Board is always open to ideas from the membership. If you see a problem with something (e.g., international relations, dealing with the media, relations with other scientists, development of research standards), you can be part of the solution by offering specific suggestions or by volunteering to serve on committees to deal with the problems about which you care the most. In addition, because of our diversity, the membership of the PA has a variety of talents, approaches, and experiences, all of which can be useful in furthering the professional goals of the Association.

There are a lot of hardworking and intelligent people out there dedicating themselves to fieldwork. They are publishing, on their websites and in popular books, tantalizing hints of data from which the field as a whole could benefit, both scientifically and methodologically. Many of these folks also seem eager for more education, for excellence: Many of them want to do the best job possible for their clients and for themselves, whatever their ideologies. What should our response be?

Since Carlos S. Alvarado’s (1989) article “The Language Barrier in Parapsychology,” English-speaking parapsychologists have been cautioned about the importance of international communication. At the same time, there has been other evidence of interest in knowing and understanding the work of non-English-speaking parapsychologists. International communication has been discussed in a number of other articles, commentaries, and conventions.

I would say that some of the language barriers that exist may be overcome if—in addition to contact through correspondence with PA members and articles published in international journals—there is increased personal contact among parapsychologists around the globe. It is obvious that occupational commitments and economic difficulties are the first obstacles that impede this contact, but frequently it is the only road to a confident and dynamic flow of communication. An example of the results of such personal contact comes from recent PA Conventions in three countries—France, Brazil, and Italy—where attenders had the opportunity to strengthen a professional association and make friendships that transcended the boundaries of geography and language.

The language barriers actually diminish, compared to the geographical barriers in parapsychology. On the other hand, a good initiative to help to demolish such barriers is the ongoing international communication between Spanish- and Portuguese-speaking parapsychologists. Empirical evidence shows more continuous communication among English-speaking parapsychologists than among Latin American parapsychologists (Alvarado, 1989). Although the problem is a difficult one to solve, some measures may be taken to implement strategies to minimize the language barrier. Since English is generally read by most parapsychologists, perhaps this language can be used to build a bridge.

Although these days it seems to be fashionable to say that we need to be more international in the membership department, I like to think I have had some positive influence in non-English-speaking countries. As president, I would recommend that we put forth special effort in raising funds for our organization. The only solution is “new blood.” We are lucky that this year there was a “new face” running for President, and that some more “new faces” ran for the Board. We need new members at every level.

Maybe the scientists should place themselves in a better position to judge reality. One can perceive the sense of malaise of some of these scientists when they realize that whatever they know about reality is actually so very poor, limited, and defective. The question underneath is: On which part of the track are we today? Some parapsychologists claim to have reached the goal. They say that the evidence is sufficient, or at least very suggestive, to affirm that psi exists (the investigation directed to proof). So the “new goal” is to explore the forms by which psi is manifested (the investigation directed to the process). Parapsychologists are also telling us that both types of persons often place nails on the racing track due to their intolerant models of reality and their habit of examining with prejudice the quantity and quality of the evidence.
Supposing that the search for psi were a racetrack where each driver is a parapsychologist in an attempt to reach the goal, the question is: What is the prize for reaching the goal? The less ambitious could say that the prize is to prove definitively the existence of psi and to banish every doubt from the minds of the skeptics and society in general. Others—maybe more ambitious—could say that the prize is the control of psi. Assuming that nobody has yet reached the goal, the pessimists could say that the runners are closer to the point of departure, and the more optimistic could say that the runners are closer or very close to the final goal.

References


Instituto de Psicología Paranormal
Salta 2015 (C1137ACQ)
Buenos Aires, Argentina
rapp@fibertel.com.ar
CAN PARAPSYCHOLOGY MOVE BEYOND THE CONTROVERSIES OF RETROSPECTIVE META-ANALYSES?

By J. E. Kennedy

ABSTRACT: Retrospective meta-analyses are post hoc analyses that have not been effective at resolving scientific controversies, particularly when based on substantially underpowered experiments. Evaluations of moderating factors, including study flaws, small-study effects, and other sources of heterogeneity, do not neutralize confounding as in a well-designed experiment and cannot fully compensate for weaknesses in the original experiments. A group of well-designed experiments with adequate power and reliable results is needed for convincing evidence for a controversial effect. The widely recommended standard for experimental research is adequate power to obtain significant results on at least 80% of confirmatory experiments. Meta-analyses in parapsychology typically have found that 20% to 33% of studies with good methodology obtained significant results. Power analysis during experimental design is needed to achieve much better replication rates. Meta-analyses of RNG studies have consistently found that z value does not increase with samples size—which is contrary to statistical theory and has been and will be interpreted as an indication of methodological problems. This anomalous property and other sources of heterogeneity for parapsychological results must be addressed. Challenging topics such as experimenter effects, goal-oriented psi, and capricious psi-missing can no longer be ignored in research syntheses.

Keywords: power analysis, meta-analysis, experimenter effects, heterogeneity, synthesis-generated evidence

The field of parapsychology remains highly controversial and has not obtained the degree of acceptance and support that is needed. For the past 25 years, meta-analyses have been the foundation for the debates about the evidence for psi. This article focuses on the questions why have the meta-analyses been controversial and what can be done to move beyond these controversies?

Although the issues described here manifest in meta-analyses, the discussion covers much more than meta-analyses. Some of the key issues originate with the methodology and findings in the original experiments and must be addressed by appropriate new experiments. Also, alternative strategies for research synthesis may avoid some of the controversies associated with meta-analysis. Most of the final recommendations here do not involve meta-analysis.

The topics covered can be categorized as (a) intrinsic limitations of meta-analysis, (b) unfortunate experimental practices in parapsychological research, (c) problematic properties of the experimental findings in parapsychology, and (d) unfortunate meta-analysis practices in parapsychology. The combination of these factors has made parapsychological meta-analyses controversial. These categories interact, which requires that the same or similar topics are sometimes discussed under multiple categories.

This article does not attempt to comprehensively discuss all aspects of every issue. Some of the topics are controversial. The purpose here is to describe enough of the differing opinions to indicate practices that are not convincing if challenged.

Intrinsic Limitations of Meta-Analyses

The advent of meta-analysis in parapsychology in the 1980s was greeted with great enthusiasm. Small studies could be integrated to provide quantitative evidence for an effect and to evaluate potential moderating factors. Rosenthal (1986) and Utts (1986, 1991) argued that effect size was a more appropriate measure of replication than statistical significance. The usual practice of ignoring power analysis when designing experiments appeared to have good justification. Large studies were not needed. Meta-analysis was considered to provide the definitive evaluation of a line of research and to provide compelling evidence for psi. Broughton (1991) described meta-analysis as a “controversy killer.”

However, this early optimism was not realized in practice. After noting cases when meta-analysis has been applied to controversial topics in psychology, Ferguson and Heene (2012) recently commented:
We have seldom seen a meta-analysis resolve a controversial debate in a field. ... We observe that the notion that meta-analyses are arbiters of data-driven debates does not appear to hold true. ... Meta-analyses may be used in such debates to essentially confound the process of replication and falsification. ... Focusing on the average effect size may be used to, in effect, brush the issue of failed replication under the theoretical rug .... (p. 558).

The controversial debates noted in the article did not include parapsychology, but the comments aptly describe the experience with meta-analysis in parapsychology.

The limitations of meta-analyses were also apparent in medical research. Inconsistent or contradictory conclusions had been reached in different meta-analyses of the same database (Bailar, 1997). The statistical book most frequently used at a pharmaceutical company I recently worked with said the following:

Our inclusion of [meta-analysis] in a chapter on exploratory analyses is an indication of our belief that the importance of meta-analysis lies mainly in exploration, not confirmation. In settling therapeutic issues, a meta-analysis is a poor substitute for one large well-conducted trial. In particular, the expectation that a meta-analysis will be done does not justify designing studies that are too small to detect realistic differences with adequate power. (Green, Benedetti, & Crowley, 2003, p. 231)

Ioannidis (2005) reached similar conclusions after developing methods for quantitative comparison of the “positive predictive value” PPV for different research methods. For adequately powered randomized experiments with little bias, he estimated the PPV to be .85. For meta-analyses of underpowered studies, the estimated PPV was .41, about half the PPV for a well-designed experiment. Similarly, evidence in pharmaceutical research is based on well-conducted experiments with adequate statistical power and reliable hypothesis tests (Food and Drug Administration, 1998). Retrospective meta-analyses cannot substitute for these well-designed, adequately powered experiments.

Several factors contribute to the limitations of meta-analyses. Some of the limitations have been discussed previously in parapsychological writings (Kennedy, 2004; Murray, 2011).

Many Choices for Post Hoc Analyses

Retrospective meta-analysis is a form of post hoc analysis. Like other types of post hoc analyses, meta-analysis involves many methodological decisions, including about statistical methods, study selection criteria, data trimming, data transformations, study quality ratings, and moderating factors. Many decisions do not have clear right and wrong answers. Different choices can result in different outcomes—which causes ambiguity and opportunity for selecting a preferred outcome.

The effects of different methodological decisions can be striking. For example, Bösch, Steinkamp, and Boller (2006a) describe some methodological differences between two meta-analyses by Radin on PK with electronic random number generators (RNGs; Radin, 1997; Radin & Nelson, 1989). The second meta-analysis reported an overall effect size that was much larger than in the previous analysis.

The increase has two sources. First, Radin removed the 258 PEAR laboratory studies included in the first meta-analysis (without discussing why), and second, he presented simple mean values instead of weighted means as presented 10 years earlier. The use of simple mean values in meta-analyses is generally discredited … because it does not reflect the more accurate estimates of effect size provided by larger studies. In the case of the data in Radin’s book, the difference between computing an overall effect size using mean values and using weighted mean values is dramatic. The removal of the PEAR laboratory studies effectively increased the impact of other small studies that had very large effect sizes. (Bösch, Steinkamp, and Boller, 2006a, p. 501)

The wide range of possible outcomes is also indicated by the RNG meta-analysis reported by Bösch, Steinkamp, and Boller (2006a). The overall outcome could be either significantly positive or significantly
negative (psi-missing), depending on whether a fixed-effects or random-effects model was used and whether the three largest studies were excluded as outliers.

As another example, the meta-analysis by Milton and Wiseman (1999) did not find significant evidence for psi, and was widely criticized for using cutoff criteria that excluded a highly significant study by Dalton (1997a). The subsequent meta-analysis by Storm, Tressoldi, and Di Risio (2010a) considered the Dalton study an outlier and excluded it from the analyses. However, other studies that had similar or larger effect sizes were not excluded.

For a typical retrospective meta-analysis, critics of the findings usually can easily find methodological decisions to challenge. These debates derive from the post hoc nature of meta-analysis and the associated potential for bias. The result has been endless controversies about meta-analysis methodology and findings in parapsychology (Bösch, Steinkamp, & Boller, 2006a, 2006b; Hyman, 2010; Radin, Nelson, Dobyns, & Houtkooper, 2006; Schmeidler & Edge, 1999; Storm, 2000; Storm, Tressoldi, & Di Risio, 2010a, 2010b). Like other types of post hoc analyses, meta-analysis can have value, but has limited effectiveness for resolving scientific controversies.

The Observational Nature of Moderating Factors

The evaluation of moderating factors in meta-analysis is observational or correlational analysis that does not neutralize confounding factors as in a well-designed experiment (Cooper & Hedges, 2009). For example, an evaluation of experimenter differences in a meta-analysis will typically be confounded by the experimenters’ testing different pools of participants in different studies. Any differences could be due to the participants rather than the experimenters. Convincing evidence for experimenter differences must be based on planned experimental comparisons with different experimenters using the same pool of participants and same testing methods. In properly designed experiments, the independent or predictor variables are manipulated to eliminate confounding. Observational data do not have these experimental controls for confounding.

Meta-analysis methodologists now recognize that convincing conclusions about causality can come only from properly designed experiments, that is, study-generated evidence rather than synthesis-generated evidence (e.g., Cooper & Hedges, 2009). Synthesis-generated evidence “help[s] ensure that the next wave of primary research is sent off in the most illuminating direction” (Cooper & Hedges, 2009, p. 564).

The observational nature of moderating factors in meta-analyses is a significant limitation. For example, an evaluation of experimental flaws is not compelling evidence about the actual effects of the flaws. These correlational analyses are synthesis-generated evidence that cannot fully compensate for poor methodology in the original experiments. Confounding can cancel or dilute a real effect as well as artificially produce an effect. Carpenter and Palmer (1998) described a detailed example of apparent confounding in a meta-analysis in parapsychology.

Heterogeneity

When heterogeneous effect sizes are found in meta-analyses, “it is unclear whether the various research findings represent a common underlying phenomenon” (Wood & Eagly, 2009, p. 459). The usual recommendation is to identify the sources of heterogeneity and use appropriate subgroups or models for the moderating factors (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009). If the sources of heterogeneity cannot be fully identified, random-effects analyses are generally considered more appropriate.

Moderating factors have a central role when working with heterogeneous data. However, the observational nature of moderating factors brings into focus the limitations of drawing conclusions from heterogeneous data. Meta-analyses should incorporate known moderating factors, but the possibility of unknown confounding factors is always present with observational data.

Most of the issues discussed in later sections of this paper typically manifest in a meta-analysis as heterogeneous effect sizes and cannot be convincingly resolved by the evaluation of moderating variables. Appropriately designed experimental studies are required.
Biases From Small Studies

In medical research it has become well established that substantially underpowered studies are prone to elevated effects known as small-study effects (Egger, Smith, Schneider, & Minder, 1997; Sterne, Egger, & Smith, 2001; Sterne, Gavaghan, & Egger, 2000). Most, but not all, factors causing small-study effects are forms of methodological bias. Substantially underpowered studies tend to be exploratory and therefore subject to practices like multiple analyses, post hoc analyses, data selection, and optional stopping. Investigators and journals tend to not report nonsignificant analyses from underpowered studies because they are inconclusive. However, significant findings are likely to be reported. Small studies usually have fewer experimental personnel and less formal procedures than large studies, and therefore have higher potential for some type of experimenter effect, including fraud. In addition, small studies are also more likely to have a limited or selected range of subjects. In general, the greater effort to plan and conduct experiments with larger sample sizes usually results in better methodology and greater likelihood of full publication with peer review.

The need to evaluate and handle small-study effects in meta-analyses is widely recognized. Various methods have been proposed (Begg, 1994; Egger, Smith, Schneider, & Minder, 1997; Ioannidis & Trikalinos, 2007b; Sutton, 2009).

However, these methods provide synthesis-generated evidence and are unreliable if the effect sizes are heterogeneous, if all the studies are small, or if there are few studies (Ioannidis and Trikalinos, 2007a, 2007b; Sutton, 2009). Meta-analyses dominated by underpowered studies can be expected to be controversial. Large studies are the frame of reference for evaluating small-study effects. The reliability and strength of the evidence increase as the number of studies with adequate power increases.

Conclusions

The original hope that retrospective meta-analysis would provide convincing evidence for a controversial effect has not been realized in practice. After describing the post hoc, observational aspects of meta-analysis, Cooper and Hedges (2009) emphasized that “a research synthesis should never be considered a replacement for new primary research” (p. 564).

The most convincing evidence for a meta-analysis occurs when all included experiments are well designed with adequate power and obtain reliable effects. If some experiments are underpowered or have flaws, or the effects are heterogeneous, observational synthesis-generated analyses are used to attempt to compensate for the weaknesses in the original experiments. The conclusions are much less convincing if weak studies are a significant part of the evidence. A group of well-designed experiments with adequate power and reliable results is needed for convincing evidence for a controversial effect.

Unfortunate Experimental Practices

The use of power analysis to design adequately powered experiments is essential for controversial areas of research. Unfortunately, power analysis has rarely been used in designing parapsychological experiments. The overly optimistic reliance on meta-analyses apparently resulted in many experimenters ignoring power analysis. However, experimenters increasingly appear to recognize that this strategy is not effective for a controversial area of research.

An underpowered study is a biased form of research because it cannot provide basic evidence that the experimental hypothesis is false. An underpowered study is likely to produce nonsignificant results and creates ambiguity about whether nonsignificant results are due to the lack of power or to the experimental hypothesis being false. A significant result is interpreted as evidence supporting the experimental hypothesis, but a nonsignificant result is inconclusive. However, for an adequately powered study, a nonsignificant result is evidence that the experimental hypothesis is false.

Underpowered studies can also generate false positive results. Bakker, van Dijik, and Wicherts (2012) discussed the common practice in academic psychology of conducting a series of small, underpowered studies rather than a large study. This strategy generates many analyses that are not treated as multiple
analyses and therefore enhance false positive results. Simulation studies verified that “executing multiple small and underpowered studies represents the optimal strategy for individual players to generate a \( p \) value of less than .05 [by capitalizing on chance]” (p. 547). This practice has also been common in parapsychology, and some examples for ganzfeld research are noted below.

In my two decades of work in medical research, the majority of researchers preferred to evaluate the evidence for an effect by focusing on adequately powered studies and discounting small, underpowered studies. This strategy has also been recommended in psychology (Kraemer, Gardner, Brooks, & Yesavage, 1998). The primary value of small studies is to justify the effort for conducting larger studies and to develop parameters for the studies.

The minimum power typically recommended in both behavioral and medical research is .8 (e.g., Cohen, 1988, p. 56; Food and Drug Administration, 1998, p. 22). If an effect is real, at least 80% of properly designed confirmatory studies should obtain significant outcomes. This degree of replication provides convincing evidence that the experimenters understand and control the phenomena being investigated. A power of .90 or .95 is preferable when possible.

As shown in Table 1, the majority of meta-analyses in parapsychology have found that 20% to 33% of the individual studies obtained statistically significant results. For meta-analyses with 30 or more studies with good methodology by a variety of experimenters, the rate of successful replication is between 20% and 33%. Cases with a smaller number of studies and/or possible methodological problems sometimes have replication rates outside of this range.

By the usual methodological standards recommended for experimental research, there have been no well-designed ganzfeld experiments. Based on available data, Rosenthal (1986), Utts (1991), and Dalton (1997b) described 33% as the expected hit rate for a typical ganzfeld experiment where 25% is expected by chance. With this hit rate, a sample size of 201 is needed to have a .8 probability of obtaining a .05 result one-tailed.1 No existing ganzfeld experiments were preplanned with that sample size. The median sample size in recent studies was 40 trials, which has a power of .22.2

Conducting a series of small studies has been a common ganzfeld research strategy. Research programs with more than 200 total trials from a series of small studies have been reported by Bem and Honorton (1994), Kanthamani and Broughton (1994), and Broughton and Alexander (1997). The combined results were significant for Bem and Honorton, but not for the other two research programs. The total numbers of trials and studies in these experimental programs apparently were not preplanned, which introduces a post hoc component for the combined analyses.

The replication rates of 20% to 33% appear to apply to all types of studies. The RNG studies are the largest experimental database and have some very large experiments, yet 25% or less were significant.

Conclusions

The most convincing evidence for an effect is driven by well-designed confirmatory experiments with adequate power to reliably obtain significant results. For the past 25 years, the efforts to develop evidence for psi have focused more on retrospective meta-analyses of underpowered studies than on conducting well-designed confirmatory experiments.

Problematic Properties of Psi Experiments

Small-Study Effects

Meta-analyses in parapsychology have often found characteristics expected for small-study effects. In the first ganzfeld meta-analysis, Hyman (1985) reported that the small studies had unexpectedly large

1 The sample size of 201 is from the free G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007; 2012) that uses an exact power calculation. A power calculation using an arcsin transformation (SWOG, n.d.) gives a sample size of 199. Most power calculators use a normal approximation, which gives a sample size of 192.

2 The power calculation method is more important for small sample sizes such as 40. The exact method gives a power of .22, the arcsin method gives .30, and the normal approximation gives .33. Exact power plotted against sample size can have a saw-tooth or jagged form for small sample sizes.
effects. Similar findings were reported for later ganzfeld studies in the Bem and Honorton (1994) meta-analysis, and in the RNG meta-analyses by Radin, May, and Thomson (1985), Radin and Nelson (2003), and Bösch, Steinkamp, and Boller (2006a). The negative relationships between scoring rate and study size reported for the early ESP studies (Nash, 1989) also indicated larger effects for small studies. In the absence of a convincing alternative explanation, cautious scientists will assume these effects result from methodological biases.

### Table 1 Properties of Meta-Analyses in Parapsychology

<table>
<thead>
<tr>
<th>Meta-Analysis</th>
<th>Significant studies</th>
<th>$N$ trials in the studies</th>
<th>$z$ &amp; $\sqrt{N}$ trials correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Counts/Percent</td>
<td>Median/Max</td>
<td>$r/p$</td>
</tr>
<tr>
<td>Ganzfeld</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honorton, 1985</td>
<td>12/28 43%</td>
<td>28/100</td>
<td>.13/.25</td>
</tr>
<tr>
<td>Bem &amp; Honorton, 1994</td>
<td>2/11 18%</td>
<td>35/50</td>
<td>-.12/.64</td>
</tr>
<tr>
<td>- Unbiased</td>
<td>1/10 10%</td>
<td>36/50</td>
<td>-.08/.59</td>
</tr>
<tr>
<td>Milton, Tressoldi, &amp; Di Risio, 2010a</td>
<td>6/30 20%</td>
<td>40/100</td>
<td>.16/.20</td>
</tr>
<tr>
<td>- All</td>
<td>10/30 33%</td>
<td>40/138</td>
<td>.48/.004</td>
</tr>
<tr>
<td>- Trimmed</td>
<td>9/29 31%</td>
<td>40/138</td>
<td>.38/.02</td>
</tr>
<tr>
<td>Combine 3 from above</td>
<td>18/71 25%</td>
<td>40/138</td>
<td>.34/.002</td>
</tr>
<tr>
<td>(1994, 1999, 2010a)</td>
<td>16/69 23%</td>
<td>40/138</td>
<td>.28/.009</td>
</tr>
<tr>
<td>Bem, Palmer &amp; Broughton, 2001a, 2001b</td>
<td>9/29 31%</td>
<td>40/128</td>
<td>.34/.04</td>
</tr>
<tr>
<td>RNG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radin &amp; Nelson, 1989</td>
<td>152/597 25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Trimmed</td>
<td>490</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radin &amp; Nelson, 2003</td>
<td>515</td>
<td>3.9x10^8</td>
<td>-.02/.36</td>
</tr>
<tr>
<td>Bösch, Steinkamp, Boller, 2006a</td>
<td>83/380 22%</td>
<td>8596/ &gt;10^8</td>
<td>-.14/.006</td>
</tr>
<tr>
<td>- Unbiased</td>
<td>83/377 22%</td>
<td>8039/ &gt;10^8</td>
<td>-.02/.66</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honorton &amp; Ferrari, 1989.</td>
<td>92/309 30%</td>
<td>1194/ 3.0x10^5</td>
<td>.16/.003</td>
</tr>
<tr>
<td>Forced-choice precognition</td>
<td>62/248 25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Trimmed</td>
<td>65/148 44%</td>
<td>5500/ 2.4x10^5</td>
<td></td>
</tr>
<tr>
<td>Radin &amp; Ferrari, 1991.</td>
<td>23/69 33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Unbiased</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PK with dice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Trimmed</td>
<td>3/16 19%</td>
<td>44/120</td>
<td>.40/.06</td>
</tr>
<tr>
<td>- Noise reduction</td>
<td>3/21 14%</td>
<td>54/937</td>
<td>-.20/.80</td>
</tr>
<tr>
<td>Lawrence, 1998. Sheep-Goat</td>
<td>18/73 24%</td>
<td>5750/ 5x10^4</td>
<td>.22/.08</td>
</tr>
<tr>
<td>Storm, Tressoldi, &amp; Di Risio, 2010a</td>
<td>3/16 19%</td>
<td>44/120</td>
<td>.40/.06</td>
</tr>
<tr>
<td>- Trimmed</td>
<td>0/14 0%</td>
<td>76/937</td>
<td>.11/.35</td>
</tr>
</tbody>
</table>

The meta-analyses shown here have been prominent and have a reasonable number and quality of studies. Other meta-analyses with few studies or with studies with questionable quality were not considered suitable for the analyses here. The counts of significant studies are all positive one-tailed results. $N$ Trials are the number of trials or individual random events in the experiment. The correlations for $z$ and $\sqrt{N}$ are one-tailed Pearson correlations unless noted otherwise. According to standard statistical theory, $z$ should increase linearly with $\sqrt{N}$.

For Bem and Honorton (1994), the original report focused on 10 studies and excluded one that had potential confounding by response bias.

For Milton and Wiseman (1999) two data values are corrected per Bem, Palmer, and Broughton (2001b).

Radin and Nelson (2003) is an update of Radin and Nelson (1989) with additional studies. The number of studies is reported as smaller (515 compared to 597) because they collapsed the 258 PEAR studies into one data point for the 2003 analysis. No rationale was given for that decision. Both meta-analyses included nonintentional and nonhuman studies and some studies with pseudo-RNGs. The $z$ and $\sqrt{N}$ correlation is two-tailed and was reported in the original meta-analysis.

For Bösch, Steinkamp, and Boller (2006a), the $z$ and $\sqrt{N}$ correlations are two-tailed and were reported in the original meta-analysis.

For Honorton and Ferrari (1989), the correlation is apparently a Pearson correlation between $z$ and $N$ (not $\sqrt{N}$) and was reported in the original meta-analysis.
A simple and intuitive way to evaluate the consistency of effects for a set of studies is to examine the relationship between \( z \) and samples size. According to standard statistical theory, the \( z \) value is expected to increase linearly with the square root of the sample size. This is the rationale for using \( z \) divided by \( \sqrt{N} \) as an effect size measure that is independent of sample size. This relationship is also the basis for power analysis. Correlations between \( z \) and \( \sqrt{N} \) for meta-analyses in parapsychology are given in Table 1. Of course, these are synthesis-generated evidence that primarily motivate future research.

The most striking finding is that the RNG studies clearly show a complete absence of positive correlation between \( z \) and \( \sqrt{N} \). The fact that \( z \) is independent of sample size has been consistently recognized from the earliest meta-analyses of RNG studies (Bösch, Steinkamp, & Boller, 2006a; Radin, May, & Thomson, 1985; Radin & Nelson, 2003; Radin, Nelson, Dobyns, & Houtkooper, 2006).

The matter is less clear for the ganzfeld studies. Consistent with the RNG studies, the first three ganzfeld meta-analyses in Table 1 did not find significantly positive (\( p \leq .1 \)) correlations. However, the meta-analysis by Storm, Tressoldi, and Di Risio (2010a) did show a significantly positive correlation between \( z \) and \( \sqrt{N} \). The positive relationship is also apparent when those data are combined with data from earlier meta-analyses as shown in Table 1.

One important question is whether the first three ganzfeld meta-analyses had sufficient statistical power to demonstrate the correlation between \( z \) and \( \sqrt{N} \). I performed simulations that assumed the overall hit rate for each meta-analysis applied for all studies in the meta-analysis (the assumptions for a fixed-effects model). The power for generating a correlation with \( p \leq .1 \) was evaluated, which is generally considered an appropriate significance level for this type of analysis. Data were generated simulating 2,000 meta-analyses for the studies in each meta-analysis. For the Honorton (1985) meta-analysis (hit rate 36.8%, based on equivalent hit rate for the \( z \) value for studies with chance not 25%), the correlation between \( z \) and \( \sqrt{N} \) had a .83 power of detecting a correlation at the \( p \leq .1 \) level. A correlation as small or smaller than the observed correlation (\( r = .13 \)) occurred on only 5% of the simulations. For the Bem and Honorton (1994) meta-analysis (hit rate 32.2%), the power for the correlation was only .33. For Milton and Wiseman (1999; hit rate 27.6% from Milton, 1999, p. 313), the power was only .24. For Storm, Tressoldi, and Di Risio (2010a; hit rate 32.2%), the power was .77.

These results indicate that a significant correlation would be expected on the Honorton (1985) and Storm, Tressoldi, and Di Risio meta-analyses, but would not be expected for the Bem and Honorton meta-analyses and for the Milton and Wiseman meta-analysis. The lack of correlation for Honorton (1985) confirms Hyman’s (1985) point that the small studies had larger effects than would be expected. Honorton (1985) argued that this result is due to differences in the experimental conditions and psychological factors rather than to methodological problems.

The meta-analysis of forced-choice precognition experiments by Honorton and Ferrari (1989) reported a significantly positive correlation, but these findings are questionable. These studies had very diverse experimental methods and extreme heterogeneity of results. The findings are questionable until the sources of heterogeneity are better understood. For example, the studies with RNGs may have different properties than the studies with cards, and these differences could confound the analyses. A favorable correlation was also reported for the meta-analysis of sheep-goat studies by Lawrence (1998), but an evaluation of the heterogeneity of the data was not reported. The other meta-analyses in Table 1 have few studies and varying results.

The issues discussed in this section can be convincingly resolved by demonstrating that future experiments with adequate power consistently obtain significant results. However, the usual methods for designing experiments do not apply if \( z \) values are unrelated to sample size. Unless the causes for this characteristic of the data are understood and controlled, RNG research will continue to have the following properties:

1. Power analysis cannot be used to design experiments. Larger sample sizes are not more likely to produce significant results. The replication rate may be limited to about 25%.
2. Standard statistical methods such as $t$ tests, binomial tests, and ANOVA assume that each trial or subject is independent of the other trials or subjects and that power analysis is applicable. However, if the $z$ value for an experiment is unrelated to sample size, these assumptions are violated in a way that makes the usual interpretations invalid.

3. The rationale for the usual measures of effect size in meta-analysis breaks down if $z$ is independent of sample size. This makes the usual meta-analysis interpretations invalid.

4. Larger effect sizes for small studies are an established symptom of methodological problems and occur when $z$ is independent of sample size. The majority of cautious scientists will find methodological problems to be the most plausible explanation for these results (e.g., Bösch, Steinkamp, & Boller, 2006a).

Several factors may have a role in these anomalous properties.

**Experimenter Effects**

Experimenter effects are one of the greatest challenges for parapsychology and should have a central place in any discussion of evidence for psi. Prominent experimenter differences have been recognized throughout the history of parapsychological research, and many experiments have established study-generated evidence for experimenter effects (see Kennedy & Taddonio, 1976; Palmer, 1997; Rao, 2011, pp. 170–197; White, 1976, for reviews). Most experimenters have often found nonsignificant outcomes on their experiments. However, a few experimenters have reported significant results on almost every experiment. The cause of the experimenter differences remains a matter of debate. The tendency for skeptics to obtain nonsignificant results is well known, but whether skepticism is more a cause or result of nonsignificant outcomes has not been resolved.

Experimenter misconduct is one possible factor contributing to the differences. Misconduct by an experimenter has occurred many times in parapsychology and is a constant threat (Kennedy, 2013). Experimenter misconduct includes biased analysis and reporting as well as fraudulently manipulating data. For example, an experimenter may present exploratory or post hoc analyses in a way that appears to be planned analyses. Or, an experimenter may conduct multiple hypothesis tests on an experiment but obtain a significant outcome on only one test, and then report it without mentioning that it was selected from multiple analyses.

Experimenter misconduct can be greatly reduced with prospective registration of studies, multiple-experimenter designs, and sharing data for independent analyses (Kennedy, 2013). I expect that these practices would significantly reduce the success rate for some experimenters.

Experimenter effects can cause apparent declines of effects. The initial findings for a line of research are usually from highly successful experimenters with 80% or more of the studies reported as significant. For example, the first six ganzfeld experiments by Honorton were all reported as significant (Honorton, 1977). Similarly, Schmidt (1973) reported that eight of the first nine studies he conducted with RNGs were significant. When other experimenters conduct replications, their rates of success are lower. The replication rates apparently decline to about 20% to 33% and then drift within this range as shown in Table 1.

**Goal-oriented psi experimenter effects.** The possibility that an experimenter influences the experimental outcome using psi is the most challenging form of experimenter effect. There is much evidence supporting this hypothesis (Kennedy & Taddonio, 1976; Palmer, 1997; Rao, 2011; White, 1976), including consistent evidence that successful experimenters are also successful subjects. Goal-oriented psi experimenter effects are conceptually the simplest form of experimenter psi, and the model most consistent with the RNG database.

Experimenter effects are one of the greatest challenges for parapsychology and should have a central place in any discussion of evidence for psi. Prominent experimenter differences have been recognized throughout the history of parapsychological research, and many experiments have established study-generated evidence for experimenter effects (see Kennedy & Taddonio, 1976; Palmer, 1997; Rao, 2011, pp. 170–197; White, 1976, for reviews). Most experimenters have often found nonsignificant outcomes on their experiments. However, a few experimenters have reported significant results on almost every experiment. The cause of the experimenter differences remains a matter of debate. The tendency for skeptics to obtain nonsignificant results is well known, but whether skepticism is more a cause or result of nonsignificant outcomes has not been resolved.

Experimenter misconduct is one possible factor contributing to the differences. Misconduct by an experimenter has occurred many times in parapsychology and is a constant threat (Kennedy, 2013). Experimenter misconduct includes biased analysis and reporting as well as fraudulently manipulating data. For example, an experimenter may present exploratory or post hoc analyses in a way that appears to be planned analyses. Or, an experimenter may conduct multiple hypothesis tests on an experiment but obtain a significant outcome on only one test, and then report it without mentioning that it was selected from multiple analyses.

Experimenter misconduct can be greatly reduced with prospective registration of studies, multiple-experimenter designs, and sharing data for independent analyses (Kennedy, 2013). I expect that these practices would significantly reduce the success rate for some experimenters.

Experimenter effects can cause apparent declines of effects. The initial findings for a line of research are usually from highly successful experimenters with 80% or more of the studies reported as significant. For example, the first six ganzfeld experiments by Honorton were all reported as significant (Honorton, 1977). Similarly, Schmidt (1973) reported that eight of the first nine studies he conducted with RNGs were significant. When other experimenters conduct replications, their rates of success are lower. The replication rates apparently decline to about 20% to 33% and then drift within this range as shown in Table 1.

**Goal-oriented psi experimenter effects.** The possibility that an experimenter influences the experimental outcome using psi is the most challenging form of experimenter effect. There is much evidence supporting this hypothesis (Kennedy & Taddonio, 1976; Palmer, 1997; Rao, 2011; White, 1976), including consistent evidence that successful experimenters are also successful subjects. Goal-oriented psi experimenter effects are conceptually the simplest form of experimenter psi, and the model most consistent with the RNG database.

Experimenters could obviously use psi to influence the outcomes of their experiments. PK research is based on the assumptions that psi is guided by the motivations of a person and can influence the outcome of a random process. Experimenters typically have high motivation for their experimental outcomes and the experiments are random processes.
Goal-oriented psi experimenter effects view the entire experiment as one random event with the probability of a hit of .05. From this perspective there is no difference between a person who wants to get a six on a die throw and a person who wants to get a significant outcome on an experiment. Both cases have motivation for the outcome of a random process.

Reviews of psi experiments have concluded that the complexity of the random process does not matter for a PK effect (Kennedy, 1978; Schmidt, 1987; Stanford, 1977). The term “goal-oriented psi” refers to the idea that PK depends on the desired outcome and is independent of the complexity of the random process. Thus, the fact that conducting an experiment is a more complex process than throwing a die would be irrelevant for a PK effect.

One of the main predictions of goal-oriented psi experimenter effects is that $z$ will be unrelated to sample size on experiments (Kennedy, 1994, 1995). The larger sample sizes increase the complexity of the process but do not alter the goal for the outcome of the experiment. The RNG meta-analyses support the hypothesis of goal-oriented psi experimenter effects.

One major implication of goal-oriented psi experimenter effects is that process-oriented research is not meaningful. Successful experimenters can obtain whatever outcomes they want.

Studies of majority-vote processes also support the hypothesis of goal-oriented psi experimenter effects (Kennedy, 1995). According to standard statistical theory, majority-vote, or multiple psi efforts for one target, could be used to enhance the reliability of psi results. However, this standard method for increasing the reliability of a signal in noise is based on the same assumptions as increasing sample size in experiments. The experimental findings consistently indicate that very different outcomes are obtained on majority-vote experiments depending on the motivations and intentions of the experimenter (Kennedy, 1995). The pattern of results is consistent with efficient goal-oriented psi experimenter effects and is not consistent with the use of majority-vote processes to enhance psi accuracy.

The hypothesis of goal-oriented psi experimenter effects that are produced with the least possible occurrence of psi is consistent with the majority-vote results and with $z$ being unrelated to sample size.

**Overly Optimistic Assumptions.** The assumptions for psi may be inconsistent with the assumptions for experimental research. Psi is assumed (a) to be guided by mental motivations, intentions, and needs, and (b) to produce effects that are not constrained by known physical parameters. However, experimental methods are based on physical parameters such as sample size, blinding, and randomization. The expectation that psi conforms to the physical parameters of experiments may be another case of overly optimistic hope. Goal-oriented psi experimenter effects could make sample size irrelevant. ESP appears to make blinding in an experiment impossible. PK by an experimenter could influence the participants’ responses as well as the random events in an experiment. The ganzfeld procedure could make a participant more susceptible to psi influence by the experimenter rather than facilitate psi by the participant.

Most discussions of psi experiments appear to be based on the implicit assumption that the experiments are somehow miraculously immune to psi by the experimenter. Detached reflection may reveal that this assumption is conspicuously inconsistent with both the basic assumptions about the nature of psi and the experience with experiments. Rao (2011, p. 184) recently observed that “the psi experimenter effect remains a deeply disgusting predicament, which few researchers were and are willing to confront.”

**Capricious Psi-Missing**

The unpredictable tendency for psi effects to be significantly opposite from the intended effect has been reported throughout the history of experimental parapsychology. Rao (1965) described the “bidirectionality” of psi that “shifts the mode of psi response from hitting to missing in a rather capricious manner” (p. 245). He described this characteristic as preventing the useful application of psi.

The RNG meta-analyses have found evidence for psi-missing, including significant missing on some large experiments. Psi-missing may contribute to the unexpected properties of the RNG studies. Meta-analyses typically are based on one-tailed analyses and will likely need to be adapted for use on at least some lines of research in parapsychology.
When psi-missing occurs at the level of the experimental outcome, two-tailed analyses and existing methodology are applicable. However, when psi-missing occurs within an experiment as described by Rao (1965, 2011, pp. 149–169), experimental design is much more challenging and the issues of post hoc and multiple analyses may be very difficult to overcome.

Several psi investigators have proposed that sporadic occurrences of psi are possible, but sustained results that provide convincing evidence and useful applications are not possible (reviewed in Kennedy, 2003). Capricious psi-missing and the associated failure to produce compellingly reliable results after many decades of research inspired this hypothesis. This hypothesis aptly summarizes the state of psi research, but requires an expanded view of scientific methodology.

Conclusions

The anomalous properties of psi experiments must be confronted and understood if progress is to be made in parapsychology. Ignoring the problematic properties in hopes that they will not be noticed or will go away is not a viable strategy for success. The ambiguities from underpowered studies tend to obscure the properties of psi experiments.

The lack of power analysis in parapsychology could be due in part to researchers recognizing that it is not applicable for psi. When I began working at the Institute for Parapsychology in 1974, the lab lore was that large studies were not more likely to obtain significant results and were very possibly less likely to be significant. I did an informal literature review at that time and found the evidence consistent with the lab lore, but, of course, there were too many confounding factors for convincing conclusions. I soon formed the impression that psi-mediated experimenter effects were a dominant factor for experiments with unselected participants, and I still find that hypothesis most consistent with the overall data. If these impressions are correct, experimental research that ignores these properties of psi will never be convincing or make scientific progress.

Unfortunate Meta-Analysis Practices in Parapsychology

Heterogeneity

As noted above, the standard recommendations for handling heterogeneity in a meta-analysis are to use appropriate subgroups or models for the moderating factors and use random-effects analyses if the moderating factors cannot be identified. However, the usual practice in parapsychology has been to trim (remove) data points that make the distribution heterogeneous and then to proceed assuming fixed-effects. The likelihood that this strategy ignores or distorts basic properties of the phenomena has rarely been considered. Among the many problems with this approach is that the trimmed studies have often been the largest studies, the ones that should have received the most attention. The meta-analysis by Honorton and Ferrari (1989) that found extreme heterogeneity and handled it by trimming 20 percent of the data is a conspicuous example of methodology that should be avoided. The only confident conclusion from that meta-analysis is that extreme heterogeneity was found.

Honorton, Radin, and others have sometimes argued that psi effects are intrinsically heterogeneous due to psychological factors (e.g., Honorton, 1985; Radin, Nelson, Dobyns, & Houtkooper, 2006). Unfortunately, most meta-analyses in parapsychology have not used methods that handle this intrinsic variability. For example, experimenter effects have not received the deserved attention in meta-analyses. Rosenthal (1986) noted that for the early ganzfeld studies the “investigators differed significantly and importantly in the average magnitude of the effects they obtained” (p. 327). However, many later meta-analyses in parapsychology have not evaluated experimenter differences.

Established sources of heterogeneity should be considered even if the test for heterogeneity is not significant. These tests have low power when the number or the sample sizes of the studies are small (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Also, exclusion of large studies because they are inconsistent with small studies is a dubious practice that should be minimized and very carefully justified if done. To the extent possible, the sources of heterogeneity should be identified and included in the analyses.
Stouffer’s Z

The foundation for meta-analysis is the quantitative evaluation of the consistency of effects in different studies. Meta-analyses in parapsychology are widely discussed as evidence for replicable psi effects. Effect size is the basis for evaluating consistency of effects and replication. The standard methods to evaluate statistical significance for meta-analyses use weighted effect size measures that incorporate the consistency of effects and the greater reliability of larger studies.

However, most parapsychological meta-analyses have determined statistical significance using Stouffer’s Z, which is based on \( p \) values rather than effect sizes and does not provide inferences about consistency or replication. Stouffer’s Z tends to obscure inconsistency among studies and is particularly vulnerable to biases from small-study effects because it gives equal weight to all studies without regard for sample size.

Stouffer’s Z is recommended only when effect sizes cannot be obtained or when a researcher is searching for any evidence of an effect without inferences about replication or effect size (Borenstein, Hedges, Higgins, & Rothstein, 2009, pp. 325–330). In the latter case, the basic purposes and methodology for meta-analysis are not applicable, and convincing evidence for a controversial effect cannot be expected. Stouffer’s Z might be useful in situations when the results are basically convincing without statistical analyses—for example, well-designed, adequately powered experiments with heterogeneous but mostly significant outcomes. It might also be useful if the experimental results are goal-oriented psi experimenter effects that make sample size and effect size irrelevant.

Recommendations

The following recommendations appear to be essential if parapsychology is to progress beyond the current state of controversy.

Improve Experimental Methodology

The first recommendation is to improve experimental methodology. These practices are essential when findings will be professionally challenged. Key points include:

1. Experimental design must include appropriate sample sizes to obtain reliable results. Low replication rates and the limitations of underpowered studies do not provide convincing evidence for a controversial effect.
2. Experiments should be prospectively registered to eliminate various potential biases, as is standard procedure for areas of medical research (Kennedy, 2013). The Koestler Parapsychology Unit (2012) at the University of Edinburgh provides a study registry, as well as information about the development of other registries.
3. Experimental procedures should make it difficult for one experimenter to intentionally or unintentionally alter data (Kennedy, 2013).
4. Raw data should be made available for independent analyses. If biased data fishing is likely, an investigator may reasonably require that the recipient register the planned analyses, including corrections for multiple analyses, prior to receiving the data (Kennedy, 2013).

Parapsychological research has generally followed the methodological practices of academic psychology and manifests methodological problems that are common in academic psychology. The need to overcome these problems is increasingly recognized—as indicated by a special issue of Perspectives on Psychological Science (Pashler & Wagonmakers, 2012, available online) that should be required reading for all social and behavioral scientists. These problems have also been discussed in other recent articles (e.g., Laws, 2013; Simmons, Nelson, & Simonsohn, 2011).

In general, the field of parapsychology would benefit from looking to medical research for experimental methodology rather than to academic psychology. Study registration is a well-established
practice in medical research, whereas psychological researchers are just beginning to develop study
registries. My experience in medical research has been that formal power analysis is standard procedure,
and the limitations of a series of substantially underpowered studies are widely recognized. My experience
in parapsychology has been that formal power analyses and concern about the limitations of underpowered
studies have been rare. The references in the previous paragraph strongly suggest that experimental
psychology and parapsychology are similar on these matters.

Understand Anomalous Properties of RNG Experiments

The second recommendation is to attempt to understand the independence of $z$ and sample size
for the RNG studies. One option is to use meta-analysis to explore moderating variables. A more powerful
option is to build simulation models that have the characteristics of the RNG database. In addition to
methodological flaws and experimenter and participant differences, challenging effects like goal-oriented
psi and capricious psi-missing will likely have a role. The database of RNG studies could be posted on
the internet to encourage investigation. Of course, the ultimate goal of these post hoc explorations is
to develop predictions for prospective research. Goal-oriented psi may be investigated with experiments
using majority-vote processes.

Based on the history of psychical research, I predict that as larger studies are conducted and more
diverse experimenters are involved, the ganzfeld and other lines of research will increasingly manifest the
anomalous properties found for the database of RNG studies.

Appropriate Methods for Research Synthesis

The third recommendation is to use appropriate methods for parapsychological research
synthesis. Unexplained heterogeneity is common and psychological factors are assumed to dominate the
experimental results. The experimenters’ motivations and goals often appear to be the most important
factor for highly successful experimenters. Methods for research synthesis need to consider the variety of
moderating factors. A random-effects model appears to be generally appropriate for formal meta-analyses
in parapsychology and should always be reported, even if as a supplemental analysis.

Research syntheses that focus on well-designed confirmatory studies provide the strongest evidence
for an effect. The simple proportion of adequately powered well-designed studies that obtained significant
results can provide strong, robust evidence for an effect. In addition, a best-evidence synthesis (Slavin, 1986,
1995) that utilizes quantitative methods but limits the database to the currently best available studies is an
increasingly recognized alternative to meta-analysis. The study selection criteria usually include a minimum
sample size. Best-evidence syntheses also describe the strengths and weaknesses of the individual studies
more than in a typical meta-analysis. Of course, best-evidence synthesis, like retrospective meta-analysis, is
a form of post hoc analysis and involves decisions that can be biased.

Novel methods may need to be developed. The available data suggest that for RNG studies $z$ might
be a more appropriate measure of effect size than $z/\sqrt{N}$. This change would fundamentally alter the basic
assumptions and methodology for analyzing data. Careful consideration of the implications is needed.
The possibility that this methodological change applies primarily for certain experimenters should also be
considered.

References

559–561.

pdf+html

Can Parapsychology Move Beyond the Controversies of Retrospective Meta-Analyses?


Broomfield, CO, USA

jek@jeksite.org

http://jeksite.org/psi.htm
EST-CE QUE LA PARAPSYCHOLOGIE PEUT SORTIR DES CONTROVERSES SUR LES META-ANALYSES RETROSPECTIVES ?

RESUME : Les méta-analyses rétrospectives sont des analyses post hoc qui n’ont pas permis de résoudre des controverses scientifiques, particulièrement lorsqu’elles sont basées sur des expérimentations avec une puissance faible. Les évaluations des facteurs modérateurs, dont les biais dans les études, les effets des petites études et autres sources d’hétérogénéité, ne permettent pas de neutraliser la confusion de la même manière qu’avec des expérimentations bien conçues, et ne compensent pas complètement les faiblesses des expérimentations originales. Un groupe d’expérimentations bien conçues avec un pouvoir adéquat et des résultats fiables est nécessaire pour obtenir des preuves convaincantes d’un effet controversé. Le standard largement recommandé pour la recherche expérimentale est un pouvoir adéquat pour obtenir des résultats significatifs dans au moins 80 % des expérimentations confirmatoires. Les méta-analyses parapsychologiques ont trouvé généralement que 20 à 33 % des études avec une bonne méthodologie obtenaient des résultats significatifs. L’analyse de puissance durant la conception de l’expérimentation est nécessaire pour parvenir à de meilleurs taux de réplication. Les méta-analyses des études avec des GNA ont trouvé de façon consistante que la valeur z n’augmente pas avec la taille des échantillons – ce qui est contraire aux théories statistiques et a été – et sera – interprété comme indiquant des problèmes méthodologiques. Cette propriété anomale et d’autres sources d’hétérogénéité pour les résultats parapsychologiques doivent être examinées. Des questions subversives telles que les effets expérimentateurs, le psi orienté vers un but et le psi-missing capricieux ne peuvent plus être ignorées plus avant dans les synthèses des recherches.

¿PUEDE LA PARAPSIKOLOGÍA SUPERAR LAS CONTROVERSÍAS DE LOS METANÁLISIS RETROSPECTIVOS?

RESUMEN: Los meta-análisis retrospectives son análisis post hoc que no han sido eficaces para resolver controversias científicas, sobre todo cuando se basan en experimentos con muy poco poder estadístico. Las evaluaciones de los factores moderadores, incluyendo defectos del estudio, efectos de estudios pequeños, y otras fuentes de heterogeneidad, no neutralizan los factores extraneos como en los experimento bien diseñados y no pueden compensar totalmente las deficiencias en los experimentos originales. Se necesita un grupo de experimentos bien diseñados con suficiente poder y resultados confiables para obtener pruebas convincentes de un efecto controversial. El estándar recomendado ampliamente para la investigación experimental es un poder adecuado para obtener resultados significativos en al menos el 80% de los experimentos confirmatorios. Típicamente los meta-análisis en parapsicología han encontrado que de 20% a 33% de los estudios con buena metodología obtuvieron resultados significativos. Es necesario hacer un análisis de poder durante el diseño experimental para lograr mucho mejores tasas de replicabilidad. Los meta-análisis de estudios con RNG han encontrado consistentemente que el valor de z no aumenta con el tamaño de las muestras, lo que es contrario a la teoría estadística y ha sido y será interpretado como una indicación de problemas metodológicos. Esta propiedad anómala y otras fuentes de heterogeneidad en los resultados parapsicológicos deben confrontarse. Temas difíciles como los efectos del experimentador, psi orientada al objetivo, y valores psi caprichosamente menores que lo que se esperaría al azar no pueden ignorarse en la síntesis de la investigación.

KANN DIE PARAPSYCHOLOGIE DIE KONTROVERSEN UM RETROSPEKTIVE METAANALYSEN ÜBERWINDEN?

ZUSAMMENFASSUNG: Retrospektive Metaanalysen sind post hoc-Analysen, mittels deren es bisher nicht gelungen ist, wissenschaftliche Kontroversen beizulegen, besonders wenn sie auf Experimenten mit erheblich
PARAPSYCHOLOGICAL EXPERIENCE AS ANOMALOUS EXPERIENCE PLUS PARANORMAL ATTRIBUTION: A QUESTIONNAIRE BASED ON A NEW APPROACH TO MEASUREMENT

By Harvey J. Irwin,* Neil Dagnall,** and Kenneth Drinwater**

ABSTRACT: When persons report a parapsychological experience, they may typically be asserting 2 occurrences: that of an anomalous or seemingly inexplicable event, and their interpretation of this event in paranormal terms. Previous studies identifying correlates of the report of parapsychological experiences may have confounded these 2 factors. The authors describe a new questionnaire which teases apart the 2 factors and report a survey which applied the new measure to the assessment of several potential correlates, namely, schizotypal tendencies, emotion-based reasoning, suspension of reality testing, and executive dysfunction. Data from a convenience sample recruited online supported the potential utility of the questionnaire, although it has yet to be demonstrated that the 2 underlying factors do have different correlates.

Keywords: parapsychological experience, anomalous experience, paranormal attribution

The fundamental objectives of this study were to devise a novel procedure for assessing potential predictors of the occurrence of parapsychological experiences and to implement this procedure in the form of a new questionnaire.

Earlier questionnaires have identified psychological correlates of the occurrence (or more pedantically, the report) of parapsychological experiences. Some routine procedures for surveying parapsychological experiences nevertheless may have theoretically significant limitations. Two specific observations in this regard will now be raised.

There is a fundamental measurement problem with one type of questionnaire item sometimes used to survey parapsychological experiences. This type of item identifies such an experience by name (e.g., “telepathy,” “a ghost”), possibly adds a formal definition, and asks respondents if they have encountered this experience. Responses then may be correlated with potential psychological predictors. This approach was used, for example, by Richards (1991) and Kennedy and Kanthamani (1995), and is still evident today. Thus, in the Anomalous Experiences Inventory (AEI, Factor 1; Gallagher, Kumar, & Pekala, 1994; Kumar & Pekala, 2001), currently one of the most widely used inventories for surveying parapsychological experiences (see also Goulding & Park, 2001), some items refer to experiences explicitly implicating paranormal phenomena (e.g., “I have had a psychic experience,” “I have seen a ghost or apparition”) or to behaviors based on an assumption of the paranormal by the respondent (“I visit fortune tellers, palm readers, tarot card readers or astrologers”). The metric problem with this type of item arises from the underacknowledged fact that some people may undergo an anomalous encounter without reconstructing it in terms of the paranormal; that is, the experience may be dismissed as mere coincidence, a misperception or illusion, and the like. With the above type of item, a conventional survey of parapsychological experiences would identify such people as nonexperients because they deny ever having experienced ESP and the like, yet an independent observer might well construe some of the respondents’ experiences as parapsychological.

A second and more common type of questionnaire item does avoid this limitation. Such items do not refer to the parapsychological experience by name; rather, they provide a noncommittal phenomenological description of the experience and ask survey participants if they have had experiences of this ilk. By way of illustration, the Anomalous Experiences Inventory includes items such as “There have been events that I dreamed about before the event occurred” and “I often know what others are feeling or thinking without them telling me.” Even people who eschew any suggestion of the paranormal may therefore be willing to concede they have had such experiences. This approach may have particular appeal to anomalistic psychologists, researchers who focus primarily on the bases of anomalous experiences rather than on the more specific investigation of hypotheses about paranormal processes.
On the other hand, while this second type of item taps the incidence of anomalous or uncanny events, it also nets people who interpret their experience as paranormal as well as people who do not. The status of the experience in the respondents’ eyes therefore is not taken into account. This represents a loss of potentially instructive information. In addition, the technique limits the interpretation that may legitimately be applied to the data. For example, on the basis of responses generated by these items, many researchers are wont to draw conclusions about the correlates of “ESP experiences” and the like, despite the fact that the items merely address an anomalous experience of unattributed origin. Thus, in this type of item perceived “paranormality” on occasion may simply revert to the investigator instead of the respondent.

Matters are further confounded when inventories, such as the Anomalous Experiences Inventory (Kumar & Pekala, 2001) and the Assessment Schedule for Altered States of Consciousness (Van Quekelberghhe, Altstotter-Gleich, & Hertwick, 1991), incorporate both of the above types of item in a single index. The psychometric issue here can be formulated in the following terms.

When a person reports a parapsychological experience, she or he may typically be asserting two occurrences: that of an anomalous or seemingly inexplicable event, and her or his interpretation of this event in paranormal terms. In a rare allusion to this issue, French and Wilson (2007) discuss the impressions of people who had been conducted on a commercial tour of a reputedly haunted site and remark that the parapsychological experient and nonexperient in this context may “differ with respect to the degree to which they experience mildly anomalous sensations such as sensing a presence, sudden changes in temperature, and dizziness, and the degree to which they opt for a paranormal interpretation of such sensations” (p. 13). Many measures of proneness to parapsychological experiences therefore may be indexing mere proneness to anomalous experiences, or the inclination to appeal to paranormal explanations of life events, or a combination of these. In other words, we often do not know if the correlates of parapsychological experiences reported in the literature are telling us something about proneness to anomalous experiences, or about willingness to make paranormal attributions, or possibly about both.

On this ground, the measure designed for the current investigation, the Survey of Anomalous Experiences (SAE; see Appendix), sought to tease apart these two components. The SAE presents respondents with a series of specific anomalous or uncanny experiences with no integral reference to possible paranormal underpinnings, and if any experience is to be acknowledged affirmatively, respondents are asked to clarify whether the experience is taken as a paranormal event or, on the other hand, as an outcome of one of various possible normal (or at least, nonparanormal) mechanisms. Under this approach one can discern for each respondent both the relative incidence of anomalous experiences and the inclination to attribute such experiences to paranormal processes. Further details of the construction of the SAE are provided in the Method section of this article.

Having designed the SAE, we planned to explore the utility of the measure by assessing several variables for their differential relationship with the two SAE components.

What factors might predict a person’s inclination to interpret life events in terms of the paranormal? Such an attribution fundamentally implicates a belief in the paranormal; indeed, the centrality of paranormal belief to a paranormal interpretation verges on tautology. The predictors of paranormal interpretation therefore may include factors understood to be the principal foundations of paranormal belief. There are many such factors (see Irwin, 2009, for a review); the ones chosen for investigation here were some pivotal cognitive processes, namely emotion-based reasoning, suspension of reality testing, and executive dysfunction.

Emotion-based reasoning, the tendency to opt for inferences that are emotionally appealing rather than logically derived, has been found to predict the intensity of paranormal belief (Irwin, Dagnall, & Drinkwater, 2012). Suspension of reality testing, the disinclination to subject inferences to rigorous rational assessment and to re-evaluation as further relevant information comes to hand, also has been shown to correlate with paranormal belief (Dagnall, Drinkwater, Parker, & Munley, 2010; Irwin, 2003, 2004). Finally, believers in the paranormal generally have been reported to exhibit defects in central executive functioning (Wain & Spinella, 2007), the operation of those components of the cognitive-processing system responsible for planning and monitoring complex goal-directed behavior. It was therefore predicted that the tendency to interpret life events in paranormal terms, as indexed by one component of the SAE, is related to emotion-based reasoning, suspension of reality testing, and executive dysfunction.
Proneness to anomalous or uncanny experiences themselves, on the other hand, may more likely stem from personality factors, notwithstanding the involvement of mediating cognitive processes. One predictor of proneness to anomalous experience may be the personality dimension of schizotypy or degrees of subclinical psychotic-like behaviors that are distributed across the general population (Claridge & Beech, 1995). Such a relationship would be expected on the ground that one of the key factorial components of schizotypy is unusual perceptual experience (e.g., Mason, Claridge, & Jackson, 1995). Indeed, there is evidence that the association between schizotypy and unusual experiences goes beyond specifically perceptual experiences (e.g., Fleck et al., 2008; Koffel & Watson, 2009; Mason & Claridge, 1999; Startup, 1999) and also encompasses a very wide range of parapsychological and other transpersonal experiences (Alvarez López, Teixeira do Carmo, & Pueyo, 2000; Byrom, 2009; Clancy, McNally, Schacter, Lenzenweger, & Pitman, 2002; Farias, Claridge, & Lalljee, 2005; B. Greyson, personal communication, “Schizotypy and NDEs,” May 8, 2012; Maltby & Day, 2002; McCreery & Claridge, 2002; Parra, 2006; Simmonds & Roe, 2000; Wolfradt, Oubaid, Straube, Bischoff, & Mischo, 1999). It was therefore predicted that proneness to anomalous experiences, as indexed by the other component of the SAE, is related to schizotypy. Note, however, that this does not necessarily impute any pathology to parapsychological experiences (see Goulding, 2004). Further, vigorous debate continues over the ontological status of schizotypy, that is, whether this construct is fully dimensional, quasidimensional (applying only to people formally diagnosed as schizotypal or schizophrenic), or taxonic (categorical). For discussions of this issue, see Lenzenweger (2010) and Rawlings, Williams, Haslam, and Claridge (2008).

At the same time, schizotypy is also a documented correlate of paranormal belief (Dagnall, Munley, Parker, & Drinkwater, 2010; Irwin & Green, 1998–1999; Thalbourne, 1985). That is, schizotypy could well be a predictor not only of proneness to anomalous experiences but also of the inclination to appeal to paranormal explanations of anomalous experiences. In part, the relationship with paranormal belief may be due to the distortions in reasoning inherent in schizotypy (Jolley, Jones, & Hemsley, 1999; Sellen, Oaksford, & Gray, 2005; Tsakanikos, 2004; Young & Mason, 2007) and more generally to deficits in executive functions (Shrira & Tsakanikos, 2004). Schizotypy is nevertheless a multidimensional domain (Mason et al., 1995). There is a possibility, therefore, that proneness to anomalous experiences and inclination to paranormal attributions are related to various facets of schizotypy.

In summary, the following predictions were formulated:

Proneness to anomalous experiences is positively related to schizotypal tendencies (Hypothesis 1).

Proneness to paranormal attributions is positively related to emotion-based reasoning (Hypothesis 2), suspension of reality testing (Hypothesis 3), executive dysfunction (Hypothesis 4), and schizotypal tendencies (Hypothesis 5).

Finally, by way of exploring the specificity of the above relationships, the following supplementary working hypotheses were devised for evaluation; given that previous empirical research has not provided unambiguous grounds for these predictions they are posited purely for comparative purposes.

Proneness to anomalous experiences is positively related to emotion-based reasoning (Hypothesis 6), suspension of reality testing (Hypothesis 7), and executive dysfunction (Hypothesis 8).

**Method**

**Participants**

A convenience sample of 203 people participated in this study. There were 49 males and 154 females. The mean age was 30.52 years (SD = 11.41) with a range of 18–75 years. Participants included undergraduates and employees from the Manchester Metropolitan University, alumni and similar associates of the university, and some volunteers from the wider community in Manchester, England.
Materials

Participants were asked to complete five questionnaires plus a few items on demographic background. The questionnaires, in order of presentation, were the Barkley Deficits in Executive Functioning Scale (BDEFS; Barkley, 2011), the Emotion-Based Reasoning subscale of the Cognitive Biases Questionnaire (EBRS; Peters et al., 2010), the Reality Testing subscale of the Inventory of Personality Organization (RT; Lenzenweger, Clarkin, Kernberg, & Foelsch, 2001), the Oxford-Liverpool Inventory of Feelings and Experiences, Short Form (O-LIFE; Mason, Linney, & Claridge, 2005), and the Survey of Anomalous Experiences.

The Barkley Deficits in Executive Functioning Scale is a 20-item self-report measure of defects in executive functioning. The BDEFS comprises five subscales, each of four items; Barkley (2011) labels the subscales Self-Management to Time (SMT), Self-Organization (SO), Self-Restraint (SR), Self-Motivation (SM), and Self-Regulation of Emotion (SRE). SMT relates to deficiencies in time management and planning. SO concerns difficulty in organizing one’s thoughts and actions (e.g., failure to think quickly in the face of unexpected events). SR taps impulsive behavior and poor inhibition of reactions to events. SM assesses lack of effort and a need for supervision at work. SRE concerns one’s inability to sustain concentration on a task and avoid distraction. Responses to each item are made on a 4-point scale (1 = Never or Rarely, to 4 = Very Often). BDEFS subscale scores are computed as the sum of responses over the respective items and thus potentially may range from 4 to 16, with higher scores representing greater executive dysfunction. The psychometric characteristics of the BDEFS are thoroughly documented (Barkley, 2011) and are generally satisfactory.

The Cognitive Biases Questionnaire (Peters et al., 2010) is a self-report measure of reasoning biases known to be associated with the formation of psychotic delusions. The Emotion-Based Reasoning subscale comprises six items indexing the inclination to draw an inference more for its emotional appeal than for its logical adequacy. For each item a short vignette is presented and the respondent is asked to choose the one of three options that best describes how they would feel about the situation. Each response is rated on a 3-point scale (1 = absence of bias; 2 = presence of bias with some qualification; and 3 = presence of bias). Scores on the EBRS are computed as the sum of ratings over the six items and therefore may potentially range from 6 to 18, with higher scores signifying stronger reliance upon emotion-based reasoning. Peters et al. (2010) report the questionnaire has good psychometric properties.

The chosen measure of reality testing was the 20-item Reality Testing subscale of the Inventory of Personality Organization (Lenzenweger et al., 2001) and measures the inclination to draw inferences without subjecting them to appropriate rigorous critical evaluation (e.g., “I can’t tell whether certain physical sensations I’m having are real, or whether I am imagining them”). Responses to the RT are made on a 5-point scale (1 = Never True, to 5 = Always True), and a total score on the scale is computed as the sum of responses over the 20 items; that is, RT scores may range from 20 to 100. The psychometric characteristics of the scale are impressive (Lenzenweger et al., 2001).

The O-LIFE, Short Form is a 43-item index of dimensional schizotypy. The questionnaire comprises four subscales labeled Unusual Experiences (12 items), Cognitive Disorganization (11 items), Introvertive Anhedonia (10 items), and Impulsive Nonconformity (10 items). The Unusual Experiences (UE) subscale, also referred to as positive schizotypy, addresses unusual perceptual experiences and thoughts. Cognitive Disorganization (CD) concerns difficulties with attention and decision-making, with thoughts tending to be disorganized or tangential. Introvertive Anhedonia (IA), also known as negative schizotypy, relates to lack of enjoyment in social contact and a tendency to be emotionally flat. Impulsive Nonconformity (IN) indexes impulsive, violent, and reckless behaviors that generally flout rules and social conventions. Respondents are required to answer either “yes” or “no” to each item, and for each scale one point is scored for each answer given in the prescribed direction; the potential range of scores on the four subscales therefore are 0–12, 0–11, 0–10, and 0–10, respectively. In addition, as the UE subscale includes a few items about transpersonal experiences which might bias the relationship with parapsychological variables (see also Thalbourne, 1985), a modified UE score (UE-M) was computed as the sum of responses to all other UE items; the potential range of UE-M scores was 0–8. Mason et al. (2005) report that the psychometric characteristics of the O-LIFE are acceptable.
The Survey of Anomalous Experiences (SAE; see Appendix) was constructed by the first author (HJI) and comprises 20 items addressing anomalous or uncanny experiences. For each item, participants are presented with an anomalous experience described without any explicit reference to its possible paranormal basis. Participants who acknowledge having had such an experience are asked to further clarify their position by stating whether they attributed their experience to a specified paranormal process or to a specified nonparanormal process. Thus, for each item addressing an anomalous experience, the respondent has three response options of the following general form: Option 1 is “yes, and I interpreted it as a (specified) paranormal experience”; Option 2 is “yes, but I interpreted it as due to (specified) normal processes”; and Option 3 is “no.” The SAE yields two scores for each participant. First, an index of proneness to anomalous experiences (PAE) was computed as the percentage of “yes” responses (i.e., Option 1 or 2 in any item) made over the 20 items; thus, this score could range from 0% to 100%. Second, each participant’s proneness to attribute anomalous experiences to paranormal phenomena was defined as the percentage of “yes” (Option 1 or 2) responses that were “yes, paranormal” (Option 1) responses. In any exceptional cases where not a single “yes” response is made (0 out of 0, an incalculable percentage ratio), an a priori decision was made to exclude these respondents from the statistical analyses of this SAE component on the ground that, given they had had no anomalous experiences, there was no basis on which to assess the degree to which they were inclined to make paranormal attributions about such experiences. For all other participants, the index for proneness to paranormal attributions (PPA) therefore could potentially range from 0% to 100%.

Procedure

The questionnaire inventory was administered as an online survey on an electronic platform known as Bristol Online Surveys (see http://www.survey.bris.ac.uk), a system to which Manchester Metropolitan University subscribes. An invitation to participate was distributed via the university’s internal e-mail system and using also a list of alumni and other associates of the university. The aim of the study was stated as the investigation of “how a person’s anomalous or seemingly inexplicable experiences relate to aspects of their everyday thought processes.” People aged at least 18 years were said to be eligible to take part, and their participation was anonymous and completely voluntary, with withdrawal from the exercise permitted at any time. The need for frankness in responding was stressed. The system automatically prevented participation more than once by the same person.

Recruitment was terminated soon after the target of 200 completions had been achieved.

Results

Preliminary Screening of Data

As noted in the Method section, a “proneness to paranormal attribution” (PPA) index cannot be computed for any participant who has had no anomalous experiences. In the present sample of 203 participants, everyone reported having had at least one such experience. Subsequent statistical analyses of PPA data therefore are based on the full sample of 203 people. The other noteworthy item of relative frequency data is that 77 of the 203 participants (38%) acknowledged at least one anomalous experience that they interpreted as having a paranormal basis. On average, the participants reported having had 7.70 of the 20 anomalous experiences and of these, 1.25 were interpreted in paranormal terms. Further comment on these findings is made in the Discussion.

A further preliminary result was that Cronbach’s alpha for the 20 SAE items was .83. This level of internal consistency is acceptable for the purposes of our study, particularly in light of the diversity of experiences surveyed by this questionnaire. In addition, an anonymous reviewer of this paper asked that we report the internal consistency (Cronbach’s α) of the PAE and PPA. This is not a straightforward request to meet, for the reason that the PAE and PPA are not separate “scales” or subsets of items but rather, higher-order metrics derived from patterns of responses over the full set of items. Nonetheless, we appreciate the spirit in which this request is made and offer the following observations. First, as noted
above, for all responses over all 20 items, Cronbach’s α for the SAE in its entirety was determined to be .83. Second, by combining response options 1 and 2 (“yes”) and considering these positive acknowledgements of anomalous experiences against the negative (“no”) responses, we can estimate that the internal consistency (Cronbach’s α) of the PAE index was .78. Given that the experiences surveyed by the SAE range from common to rare, both of these indices of internal consistency may be deemed satisfactory. A similar estimate for the PPA, however, could not be computed: here Option 3 (“no”) has to be recoded as missing data, and as several items had no Option 1 (“yes, paranormal”) responses and thus had zero variance, any calculation of Cronbach’s α based on all 20 items was not possible.

Descriptive Statistics

Descriptive statistics for the principal variables of the study are given in Table 1. The distribution of all but two of the variables was significantly skewed; bivariate relationships therefore were indexed by Spearman correlation coefficients.

Table 1

Descriptive Statistics and Spearman Correlations Between SAE Components and Independent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Skewness</th>
<th>PAE Uncorrected</th>
<th>PPA Uncorrected</th>
<th>PAE</th>
<th>PPA</th>
<th>Uncorrected p</th>
<th>PAE</th>
<th>Uncorrected p</th>
<th>PPA</th>
<th>Uncorrected p</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAE components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAE</td>
<td>38.47</td>
<td>18.57</td>
<td>5–100</td>
<td>.67*</td>
<td>.35***</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPA</td>
<td>13.30</td>
<td>22.38</td>
<td>0–100</td>
<td>1.86*</td>
<td>.35***</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizotypy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UE</td>
<td>3.30</td>
<td>2.81</td>
<td>0–12</td>
<td>.94*</td>
<td>.52***</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td>.41***</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UE-M</td>
<td>2.65</td>
<td>2.12</td>
<td>0–8</td>
<td>.58*</td>
<td>.50***</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td>.32***</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>5.13</td>
<td>3.14</td>
<td>0–11</td>
<td>.24</td>
<td>.29***</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
<td>.089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>2.19</td>
<td>2.13</td>
<td>0–10</td>
<td>1.09*</td>
<td>.06</td>
<td>.41</td>
<td>.04</td>
<td></td>
<td>.547</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>3.48</td>
<td>2.29</td>
<td>0–9</td>
<td>.38*</td>
<td>.27***</td>
<td>&lt;.01</td>
<td>.08</td>
<td></td>
<td>.247</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion-based reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBRS</td>
<td>7.87</td>
<td>1.93</td>
<td>6–15</td>
<td>1.16*</td>
<td>.33***</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td>.41***</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspension of reality testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>37.78</td>
<td>11.64</td>
<td>20–87</td>
<td>1.09*</td>
<td>.59***</td>
<td>&lt;.01</td>
<td></td>
<td></td>
<td>.41***</td>
<td></td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executive dysfunction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMT</td>
<td>9.48</td>
<td>2.92</td>
<td>4–16</td>
<td>.23</td>
<td>.09</td>
<td>.201</td>
<td>.04</td>
<td></td>
<td>.607</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO</td>
<td>7.20</td>
<td>2.65</td>
<td>4–16</td>
<td>.91*</td>
<td>.15</td>
<td>.028</td>
<td>.07</td>
<td></td>
<td>.299</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>6.75</td>
<td>2.44</td>
<td>4–16</td>
<td>.99*</td>
<td>.23**</td>
<td>.001</td>
<td>.05</td>
<td>.493</td>
<td>.201</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td>6.63</td>
<td>2.71</td>
<td>4–16</td>
<td>1.18*</td>
<td>.14</td>
<td>.041</td>
<td>.09</td>
<td>.201</td>
<td>.483</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRE</td>
<td>7.73</td>
<td>3.29</td>
<td>4–16</td>
<td>.90*</td>
<td>.18*</td>
<td>.011</td>
<td>.05</td>
<td></td>
<td>.483</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01, ***p < .001 (corrected within each hypothesis)

Inferential Analyses

The SAE components’ relationships with the basic demographic factors of age and gender were inspected, and the only noteworthy finding was a small correlation between PPA and age (rho = .17, corrected p < .05). Apart from its intrinsic interest, this finding indicates a need to take some account of the factor of age in subsequent analyses.

Table 1 presents the Spearman correlations between components of the SAE (PAE, PPA) and the predictor variables of schizotypy (UE and UE-M, CD, IA, and IN), emotion-based reasoning (EBRS), suspension of reality testing (RT), and executive dysfunction (SMT, SO, SR, SM, and SRE). Bonferroni corrections were applied to the significance levels of the correlation coefficients on a hypothesis-
by-hypothesis basis (Abramson et al., 1999; Shaffer, 1995). To assess the study’s stated hypotheses, the authors had planned to undertake a multiple regression analysis for each hypothesis. The common data transformations (logarithm, square root, and inverse) and the elimination of multivariate outliers failed to correct fully the skewed distribution of several of the variables, so for each hypothesis, multiple regression analysis with bootstrapping was undertaken (1,000 samples with bias corrected and accelerated analyses). Bootstrapping is a procedure for using the original sample data to estimate a variable’s distribution in the population and thereby avoids the need to meet the statistical requirement for a normal distribution (IBM Corporation, 2011). The analyses are now reported for each of the study’s hypotheses in turn.

Hypothesis 1, on the prediction of proneness to anomalous experiences by schizotypal tendencies, is supported by the significant positive correlations between PAE and all facets of schizotypy except Introvertive Anhedonia (Table 1). The hypothesis was assessed further by means of a stepwise multiple regression analysis, with bootstrapping, of PAE scores on UE, CD, IA, IN, and age. The regression equation was significant, \( F(5,197) = 11.20, p < .001 \), adjusted \( R^2 = .20 \), with UE making an independently significant contribution to the regression, partial \( r(196) = .39, p < .001 \), beta = .44. These findings were replicated even when paranormally oriented items were omitted from the UE scale, \( F(5,197) = 9.62, p < .001 \), adjusted \( R^2(196) = .18, p < .001 \). UE-M: partial \( r(196) = .35, p < .001 \), beta = .42. Taken in conjunction with the Spearman correlations (Table 1), these findings confirm Hypothesis 1.

Hypothesis 2, on the positive relationship between proneness to paranormal attributions and emotion-based reasoning, is supported by the significant correlation between PPA and EBRS scores, \( r(202) = .41, p < .001 \). The hypothesis was tested also with a stepwise multiple regression analysis, with bootstrapping, of PPA scores on EBR and age. The regression equation was significant, \( F(2,200) = 27.94, p < .001 \), adjusted \( R^2 = .22 \), with independently significant contributions to the regression made by both EBR, partial \( r(199) = .41, p < .001 \), beta = .40; and age, partial \( r = .24, p < .001 \), beta = .25. Hypothesis 2 therefore is confirmed.

Hypothesis 3, on the positive relationship between proneness to paranormal attributions and suspension of reality testing, is supported by the significant correlation between PPA and RT scores, \( r(202) = .41, p < .001 \). The hypothesis was further evaluated with a stepwise multiple regression analysis, with bootstrapping, of PPA scores on RT and age. The regression equation was significant, \( F(2,200) = 36.22, p < .001 \), adjusted \( R^2 = .26 \), with independently significant contributions to the regression made by both RT, partial \( r(199) = .47, p < .001 \), beta = .46; and age, partial \( r(199) = .31, p < .001 \), beta = .28. Hypothesis 3 is confirmed.

Hypothesis 4 concerns the prediction of proneness to paranormal attributions by executive dysfunction. All correlations between PPA and the five aspects of executive dysfunction were nonsignificant (see Table 1). The hypothesis was tested also with a stepwise multiple regression analysis, with bootstrapping, of PPA scores on SMT, SO, SR, SM, SRE, and age. The regression equation reached significance, \( F(6,196) = 3.51, p < .01 \), adjusted \( R^2 = .07 \); but only the factor of age was independently significant, partial \( r(195) = .21, p < .01 \), beta = .21. This finding, reinforced by the nonsignificant Spearman correlations, leads to the conclusion that Hypothesis 4 is not supported.

Hypothesis 5, on the prediction of proneness to paranormal attributions by schizotypal tendencies, is supported by the significant positive correlations of PPA with the two measures of unusual experiences. For UE, \( r(202) = .41, p < .001 \); for UE-M, \( r(202) = .32 \). However, other indices of schizotypal tendencies had no relationship to PPA (see Table 1). Hypothesis 5 was then assessed with a stepwise multiple regression analysis, with bootstrapping, of PPA scores on UE, CD, IA, IN, and age. The regression equation was significant, \( F(5,197) = 16.86, p < .001 \), adjusted \( R^2 = .28 \), with independently significant contributions to the regression made by UE, partial \( r(196) = .48, p < .001 \), beta = .55; and age, partial \( r(196) = .24, p < .01 \), beta = .22. These findings were replicated even when paranormally oriented items were omitted from the UE scale, \( F(5,197) = 8.69, p < .001 \), adjusted \( R^2 = .16 \). UE-M: partial \( r(196) = .32, p < .01 \), beta = .38; age, partial \( r = .25, p < .01 \), beta = .24. These findings, in conjunction with the Spearman correlations shown in Table 1, confirm Hypothesis 5 insofar as the “unusual experiences” factor or positive schizotypy is concerned.

The remainder of the hypotheses were formulated to test the specificity of the previous predictions. The following results therefore should be regarded merely as supplementary. Under Hypothesis 6, a positive relationship was posited between proneness to anomalous experiences and emotion-based reasoning. This relationship is supported by the correlation between PAE and EBRS scores, \( r(202) = .33, p < .001 \), and it
was evaluated further through a stepwise multiple regression analysis, with bootstrapping, of PAE scores on EBR and age. The regression equation was significant, \( F(2, 200) = 13.32, p < .001; \) adjusted \( R^2 = .12, \) with EBR making an independently significant contribution to the regression, \( r(199) = .34, p < .001, \) beta = .34. Hypothesis 6 is supported.

Hypothesis 7, on the positive relationship between proneness to anomalous experiences and suspension of reality testing, is supported by the significant correlation between PAE and RT scores, \( r(202) = .59, p < .001. \) The relationship then was tested by means of a stepwise multiple regression analysis, with bootstrapping, of PAE scores on RT and age. The regression equation was significant, \( F(2, 200) = 44.78, p < .001, \) adjusted \( R^2 = .30, \) with RT making an independently significant contribution to the regression, partial \( r(199) = .55, p < .001, \) beta = .56. Hypothesis 7 is supported.

Hypothesis 8, concerning the positive relationship between proneness to anomalous experiences and executive dysfunction, is supported by significant correlations between PAE and two facets of executive dysfunction, Self-Restraint (\( .23 \)) and Self-Regulation of Emotion, \( r(202) = .18, p < .01. \) The relationship then was assessed with a stepwise multiple regression analysis, with bootstrapping, of PAE scores on SMT, SO, SR, SM, SRE, and age. The regression equation was significant, \( F(6, 196) = 2.38, p < .05, \) adjusted \( R^2 = .04, \) with SR making an independently significant contribution to the regression, partial \( r(195) = .16, p < .05, \) beta = .19. Hypothesis 8 is supported, at least in regard to Self-Restraint.

**Discussion**

In discussing the findings of the study, it is useful to begin with a consideration of the data on the SAE itself. As indexed by this questionnaire, every single participant had had at least one anomalous experience. This result invites comparison with the level of acknowledgment in previous surveys of parapsychological experiences that are analogous to the anomalous experiences tapped here. In previous polls (e.g., Kennedy, Kanthamani, & Palmer, 1994) it is commonly reported that up to 40% of participants deny having had any parapsychological experiences at all. The present result strongly suggests that such respondents are intent on rejecting any paranormal interpretation of their life experiences rather than being immune to the occurrence of anomalous experiences per se. In this respect the SAE’s separation of the occurrence and the interpretation of experiences appears to have merit.

On the other hand, the relative incidence of participants willing to attribute at least one of their anomalous experiences to paranormal factors, at 38%, seems relatively low in this study. This figure is well below the majority of people in previous polls (e.g., Palmer, 1979) who admit to having had one or more parapsychological experiences. Perhaps in completing the SAE, participants’ attention is drawn (for the first time?) to the possibility of interpreting their anomalous experiences in nonparanormal terms, and this context induces them to be more conservative in their paranormal attributions. This interpretation is consistent with the observation that although the participants acknowledged an average 7.70 anomalous experiences, only 1.25 of these were conceded to be paranormal in nature. Future research with the SAE could explore as a contrast an item format comprising an initial query about the nominated experience, and then if the answer is affirmative, a probe to ascertain whether the experience was interpreted as having a paranormal basis (without any explicit reference to potential nonparanormal attributions).

The correlation between the two components of the SAE (\( \rho = .35 \)) was to be expected if the occurrence of anomalous experiences is one factor that facilitates a person’s inclination to make paranormal attributions of life events (e.g., see Coll & Taylor, 2004). At the same time, the relationship is not so strong as to call into question the proposition that proneness to anomalous experiences is discriminable from proneness to paranormal attributions.

The inclusion in the study of measures of various independent variables had two objectives. First, there was a need to determine if previously documented correlates of paranormal belief and parapsychological experiences could be replicated with the newly designed SAE questionnaire. Broadly speaking, the replications were successful. One or both of the components of the SAE have been shown to be significantly related to the factors of schizotypy, emotion-based reasoning, suspension of reality testing, and executive dysfunction. In this regard, the SAE’s approach to surveying parapsychological experiences is supported and its future use in this context is encouraged.
A second objective was to determine if previously documented correlates of parapsychological experiences may have been confounded by a failure to give due cognizance to a distinction between proneness to anomalous experiences and proneness to paranormal attributions to such experiences. With the present selection of independent variables, our study largely did not succeed in establishing that the two components of the SAE had a different pattern of correlates. Thus, both PAE and PPA alike were related to the factors of schizotypy, emotion-based reasoning, and suspension of reality testing. Only for the variable of executive dysfunction did the SAE components exhibit marked disparity, and this was in the opposite direction to our expectations. There is a need for further research to examine other reported correlates of parapsychological experiences to see if they actually reflect proneness to anomalous experiences, or proneness to paranormal attributions, or both.

The basis of the contrary findings for executive dysfunction is unclear. Previous research (Wain & Spinella, 2007) had identified this factor as a correlate of paranormal belief, so its failure here to predict proneness to paranormal attributions was unexpected. Perhaps even more surprising to us was the observation of a significant relationship between executive dysfunction and proneness to anomalous experiences. In an earlier study by Spinelli, Reid, and Norvilitis (2002), executive dysfunction was not a significant predictor in a multiple regression with measures of personality boundaries and gender roles on “experience with the paranormal.” On further reflection, it appears that their measure of personality boundaries in part encompasses proneness to altered states of consciousness (e.g., see Hicks, Bautista, & Hicks, 1999), and this may have been sufficient to suppress the further contribution of executive dysfunction in the analysis conducted by Spinelli et al. In following up this possibility, the first author (HJI) was able to establish that in the data of Spinelli et al. the simple bivariate correlation between parapsychological experience and (defective) executive functioning was significant ($r = .16, p < .05$; Jill M. Norvilitis, personal communication, April 11, 2012). At least, this admittedly weak trend suggests that our finding of an association with proneness to anomalous experiences is not entirely aberrant. Nonetheless, the contrary results for the two SAE components remain to be reconciled. Further investigation is currently being undertaken into the relationship between executive dysfunction and the respective SAE scales.

Another anomalous finding was a positive correlation between proneness to paranormal attributions and age. Although only weak ($\rho = .17$) this result also was contrary to expectations. By and large, demographic factors are not effective predictors of paranormal belief, but where a significant relationship with age has been reported, its direction almost always is negative (see Irwin, 2009 for a review). It is unclear why younger participants in our sample were less likely to attribute their anomalous experiences to paranormal processes. Perhaps they were more susceptible than older participants to the suggestion under the second response option that anomalous experiences could well be explicable in nonparanormal terms.

An explicitly anticipated finding was that schizotypy is a predictor of both proneness to anomalous experiences and proneness to paranormal attributions. In this regard, it is interesting that Sumich et al. (2008) report evidence to suggest these tendencies may both be facets of (positive) schizotypy but may spring from distinctly different neurophysiological processes. Perhaps the work of Sumich et al. represents independent support for the validity of the SAE’s differentiation of proneness to anomalous experiences and proneness to paranormal attributions.

Some limitations of the study must be acknowledged. One limitation concerns the representativeness of a sample recruited within an English university community through an online appeal. It is not clear, however, why the observed relationships may differ in a wider sample recruited by more traditional methods. Indeed, there is evidence (e.g., Göritz & Schumacher, 2000) that while online recruitment may yield a sample with slightly elevated paranormal beliefs, the relationships between these beliefs and various predictor variables do not differ from those found under other recruitment procedures. Additionally, on the first few occasions, the second author (ND) used online sampling in a study in which he compared scores on paper and computer versions of the questionnaires and found no substantial differences. The generality of our present findings nevertheless remains open to independent scrutiny.

A second limitation is that the order of questionnaires in the test inventory was not counterbalanced across participants. The authors elected to present the SAE as the final questionnaire for all participants in
order that responses to measures of the independent variables would not be constrained in light of what a respondent had disclosed on items about parapsychological experiences. Independent investigation with a different order of presentation nevertheless would be appropriate.

In conclusion, the construction of the SAE was undertaken in response to our perception that previously identified correlates of parapsychological experiences may be compromised by a confounding of two distinct factors, proneness to anomalous experiences and proneness to paranormal attributions. With the set of independent variables used in our study, we were not able to show that previously identified correlates of paranormal belief and parapsychological experiences relate differently to these two factors. On the other hand, our findings more generally do support the use of the SAE in future studies of the correlates of parapsychological experiences. The SAE may effectively be applied in its entirety either in its present form or with a modification in format noted earlier in the Discussion, or the style of SAE items may be applied to research into a single type of parapsychological experience.

References


*Psychology Division
School of Behavioural, Cognitive and Social Sciences
University of New England
Armidale, NSW 2351, Australia
hirwin2@une.edu.au

**Manchester Metropolitan University
Faculty of Health, Psychology, and Social Care
Department of Psychology
Gaskell Campus, Hathersage Road
Manchester, M13 0J A, UK

Abstracts in Other Languages

French

L’EXPERIENCE PARAPSYCHOLOGIQUE EN TANT QU’EXPERIENCE ANOMALE ASSOCIEE A UNE ATTRIBUTION PARANORMALE : UN QUESTIONNAIRE BASE SUR UNE NOUVELLE APPROCHE DE LA MEASURE

RESUME : Lorsque des personnes relatent une expérience parapsychologique, elles produisent en fait deux affirmations : celle d’avoir vécu un événement anomal ou apparemment inexplicable, et leur interprétation de cet événement en termes de paranormal. Les études antérieures identifiant les corrélats des expériences parapsychologiques peuvent avoir confondu ces deux facteurs. Les auteurs décrivent un nouveau questionnaire qui distinguent ces deux facteurs et présentent un sondage testant cette nouvelle mesure dans ses corrélats potentiels avec les tendances schizotypiques, le raisonnement basé sur l’émotion, la suspension du test de la réalité, et la
dysfonction exécutive. Les données d’un échantillon de convenance recruté en ligne supportent l’utilité potentielle du questionnaire, bien qu’il soit encore nécessaire de démontrer que les deux facteurs sous-jacents peuvent avoir des corrélats différents.

Spanish

EXPERIENCIAS PARAPSICOLÓGICAS COMO EXPERIENCIAS ANÓMALAS MÁS ATRIBUCIONES PARANORMALES: UN CUESTIONARIO BASADO EN UN NUEVO ENFOQUE DE MEDICIÓN

RESUMEN: Cuando las personas reportan una experiencia parapsicológica, típicamente afirman 2 eventos: un evento anómalo o aparentemente inexplicable, y la interpretación de este evento en términos paranormales. Estudios anteriores que identificaron correlatos de las experiencias parapsicológicas pueden haber confundido estos 2 factores. Los autores describen un nuevo cuestionario que escinde los 2 factores e informan de una encuesta que usó la nueva medida en la evaluación de varios correlatos posibles, es decir, tendencias esquizotípicas, razonamiento basado en la emoción, suspensión de la prueba de realidad, y disfunción ejecutiva. Los datos obtenidos con una muestra de conveniencia reclutada en la red apoya la posible utilidad del cuestionario, aunque todavía no se ha demostrado que los 2 factores subyacentes tengan diferentes correlatos.

German

PARAPSYCHOLOGISCHE ERFAHRUNG ALS ANOMALE ERFAHRUNG PLUS PARANORMALE ZUSCHREIBUNG: EIN FRAGEBOGEN AUF GRUNDLAGE EINES NEUEN MESSEVERFAHRENS


APPENDIX

Survey of Anomalous Experiences

Surveys suggest that anomalous, uncanny or seemingly inexplicable experiences are very common in our society. This questionnaire asks if you personally have had some of these experiences, and asks also what you make of them. For each item simply click on the option that most closely represents your own position; if you’ve have had the experience more than once, choose the option that is most often the case.

Q1. I have had a dream about something of which I was previously unaware, and subsequently the dream turned out to be accurate.
   Yes, and it must have been an instance of telepathy or ESP
   Yes, but it was probably just a coincidence or unwitting insight
   No

Q2. I have stared at the back of someone’s head and eventually they turned around and looked at me.
   Yes, and it must have been an instance of telepathy or ESP
   Yes, but it was probably just a coincidence or something else I did
   No
Q3. Sometimes I’ve been thinking of a person I haven’t heard from in ages, and later in the day I received a phone call, email or letter from that very person.
   Yes, and it must have been an instance of telepathy or ESP
   Yes, but it was probably just a coincidence or rational expectation
   No

Q4. With someone I know intimately I sometimes know what they are about to say before they say it.
   Yes, and it must have been an instance of telepathy or ESP
   Yes, but it was probably just a lucky guess based on my familiarity with them
   No

Q5. On at least one occasion I’ve had the impression of a figure nearby, yet nobody could possibly have been there.
   Yes, and it must have been an experience of an apparition or ghost
   Yes, but it was probably just an illusion or misperception
   No

Q6. I have become aware of a scent in a room, yet there was nothing there that could have that smell.
   Yes, and it must have been an instance of an apparition or ESP
   Yes, but it was probably just an illusion or physiological anomaly
   No

Q7. I have had an impression that a specific event was occurring at some distant location and subsequently the impression turned out to have been accurate.
   Yes, and it must have been an instance of clairvoyance or ESP
   Yes, but it was probably just a coincidence or rational expectation
   No

Q8. I have seen an envelope of light around a person, and the color of the light depended on that person’s mood or wellbeing.
   Yes, and it must have been an instance paranormal aura perception
   Yes, but it was probably just an illusion or physiological anomaly in me
   No

Q9. I have accurately foretold a future event when I could not possibly have known it would occur.
   Yes, and it must have been a case of a premonition or ESP
   Yes, but it was probably just good judgment or a coincidence
   No

Q10. I have seen a pet become excited shortly before its owner arrived back home.
   Yes, and it must have been an instance of telepathy or ESP
   Yes, but it was probably just the pet having learned when its owner would return or using its acute hearing to detect the owner’s approach
   No

Q11. On at least one occasion I have had the impression that I, my perceiving self, was outside my physical body and seeing the vicinity from an external vantage point.
   Yes, and it must have been a paranormal separation of mind from body
   Yes, but it was probably just an illusion or misperception
   No

Q12. On at least one occasion I have had the impression I was in direct contact with the spirit of a deceased person.
   Yes, and it must have been an instance of channeling or paranormal communication with a discarnate being
   Yes, but it was probably just an illusion or wishful fantasy
   No

Q13. I have had the experience of being healed by another person using only the power of their mind.
   Yes, and it must have been a case of psychic healing
   Yes, but it was probably just an effect of suggestibility
   No

Q14. On at least one occasion an object near me unaccountably moved or fell at the very time a loved one was undergoing a trauma at a distant location.
   Yes, and it must have been an example of paranormal action or psychokinesis
   Yes, but it was probably just a coincidence or a minor earth tremor
   No

Q15. I have seen (in person or on television) a psychic levitate an object.
   Yes, and it must have been an instance of paranormal action or psychokinesis
   Yes, but it was probably just a conjurer’s trick
   No
Q16. In a life-threatening situation I have had the impression that my disembodied “self” was moving along a tunnel toward a light.

Yes, and it must have been an instance of spiritual transfer to an after-death world
Yes, but it was probably just an illusion, perhaps induced by sudden physiological changes
No

Q17. When I was a child I thought I had lived as a different person in another time and place.

Yes, and it must have been an instance of reincarnation
Yes, but it was probably just an illusion or wishful fantasy
No

Q18. I have inherent abilities that neither of my (biological) parents possessed.

Yes, and these abilities I must have possessed in a previous lifetime or incarnation
Yes, probably because my life experience has differed from that of my parents
No, or don’t know

Q19. While alone in bed at night I have felt someone or something touch me, but when I switched on the light there was nobody else there.

Yes, and it must have been an instance of a ghost or a demon
Yes, but it was probably just an illusion or dream, perhaps caused by anxiety
No

Q20. In magazines I read, the horoscope for my star sign usually turns out to be accurate.

Yes, because astronomical phenomena have paranormal influences on human life
Yes, but astrologers’ statements are often true of anyone, regardless of star sign
No, or don’t know
ABSTRACT: A questionnaire, administered to a predominantly African-American sample in Northeastern North Carolina (N = 965), surveyed incidence of anomalous experience, psychological symptoms, psychological variables related to shamanism, scales pertaining to psychological well-being, and demographic variables. Multidimensional scaling analysis allowed evaluation of hypotheses drawn from sheep, goat, and “black-sheep” theories. Sheep theories predict that paranormal experiences provide direct survival advantages, derived from psi. Goat theories argue that psi does not exist; anomalous experiences are associated with psychopathology and provide no direct evolutionary benefit. A “black sheep” theory has sheep and goat elements; it portrays a psychosis-spirituality continuum, with benefits derived from spirituality; psi may exist but does not provide direct evolutionary benefits sufficient to overcome psychopathological costs. Within the black sheep paradigm, the ritual healing theory argues that shamanic variables and associated genotypes facilitated coping skills and hypnotic/placebo effects. Study findings provide mixed support for sheep and goat hypotheses but fully support black sheep hypotheses.

Keywords: dissociation, psychosis, schizotypy, hypnosis, ritual healing, shamanism

Anomalous experiences have been defined as uncommon or irregular events “believed to deviate from ordinary experience or from the usually accepted explanations of reality” (Cardeña, Lynn, & Krippner, 2000, p. 4). Such episodes include extrasensory perceptions, psychokinesis, out-of-body experiences, near-death experiences, and anomalous healing.

Research indicates a relationship between anomalous experience and psychosis, with schizotypy as a mediating variable (Clarke, 2010; Dagnall, Munley, Parker, & Drinkwater, 2010; Kelley, 2011; McClenon, 2011, 2012; McCreery & Claridge, 2002; Schumaker, 1990, 1995). Schizotypy is a psychological construct describing a continuum of personality characteristics and experiences related to psychosis, and in particular schizophrenia. Schizotypy has been defined as consisting of unusual experiences, cognitive disorganization, introverted anhedonia, and impulsive nonconformity (Chapman, Chapman, & Kwapi, 1995; Claridge, 1997; Claridge et al., 1996; Claridge & Beech, 1975; Goulding, 2004). Although some researchers argue that schizotypy is not a unified, homogeneous concept, evidence indicates that it has the same genetic basis as schizophrenia and is predictive, to a degree, of schizophrenia onset (Chapman & Chapman, 1987; Chapman, Chapman, Kwapi, Eckblad, & Zinser, 1994; Jang, Woodward, Lang, Honer, & Livesley, 2005; Lenzneither, 2010). Variables such as magical ideation and unusual experience (a category that includes psi experiences) are particularly correlated with onset probability (Chapman et al., 1994).

The ongoing incidence of psychosis is regarded as an evolutionary paradox since the disorder has genetic basis coupled with survival disadvantages (Brüne, 2004; Keller & Miller, 2006; McClenon, 2011; Pearlson & Folley, 2008; Polimeni & Reiss, 2003). How can schizophrenia genes, conferring disadvantages, persist in all societies over many centuries? Although mainstream evolutionary theories ignore paranormal experiences, explanations linking psychosis with psi have been shaped by attitudes toward the paranormal: (a) Sheep theories argue that psi experiences provide survival advantages beyond placebo effects, sufficient to overcome the evolutionary disadvantages of psychosis. (b) Goat theories deny the existence of psi. They argue that religion/spirituality is delusional and that psychopathological processes explain paranormal experiences. (c) Black sheep theories have elements drawn from both sheep and goat theories. They argue that psi has insufficient benefits to have evolutionary impact but anomalous experiences have shaped spirituality and religiosity, which confer survival advantages.

Sheep Theories

Sheep theories assume that some percentage of anomalous experiences involve psi. These episodes affect reality or provide information contributing to experiencers’ survival and fertility (Kelley, 2011).
Decades of parapsychological research, coupled with field studies, indicate that psi is real and sometimes beneficial. Twin studies indicate that ESP has genetic basis (Nash & Buzby, 1965). Cross-cultural studies of waking ESP, paranormal dreams, apparitions, out-of-body experiences, and spiritual healing reveal structural commonalities, implying genetic basis (McClennon, 1997a, 1997b, 2000, 2002a, 2002b). Evidence derived from analysis of 130 Scottish families implies that “second sight” is hereditary (Cohn, 1999). Folklore and anthropological accounts from all over the world describe paranormal benefits from magical, spiritual, and religious practices (Long, 1977; Thompson, 1955–1958).

Within the evolutionary psychology literature, the sheep argument would be considered a balance theory. Balance theorists argue that psychosis alleles confer advantages sufficient to counteract costs; some suggest that creativity and shamanism, linked to psychosis, provide these benefits (Nelson & Rawlings, 2010; Nettle, 2001; 2006; Polimeni & Riess, 2003; Stevens & Price, 2000). If psychosis is coupled with psi, the paranormal benefits could be sufficient to overcome the costs of psychosis (Kelley, 2011). Although balance theory is subject to criticism (Keller & Miller, 2006), schizophrenia alleles, whose functions are not yet known, have increased in frequency in the past, a finding that implies evolutionary benefits (Crispi, Summer, & Dorus, 2007).

Community survey findings support sheep arguments by showing relationships between anomalous experience and psychological well-being (Greeley, 1975, 1987; Kennedy & Kanthamani, 1995; Palmer & Braud, 2002). An emerging psychotherapy paradigm focuses on links between spirituality and mental well-being. Studies indicate that schizotypal experiences, when coupled with cognitive organization and adaptive religious beliefs, contribute to psychological health (Boden & Berenbaum, 2004; Goulding, 2004, 2005; McCreery & Claridge, 1995, 1996, 2002; Nelson & Rawlings, 2010; Schofield & Claridge, 2007). According to Darwin’s paradigm, evolutionary benefits tend to generate positive affect and psychological health whereas disadvantages produce negative affect and psychopathology. As a result, sheep theorists would argue that psi tends to be linked to psychological health.

Freud (1933) and Palmer (2011) provide alternate, parallel sheep theories. They suggest that telepathy could be a primitive means of communication replaced by language, one still available to people regressed to preverbal states. If this theory is true, ESP alleles should exist in humans but be more prevalent in certain animals. This theory has a “black sheep” quality in that it implies that psi does not provide many benefits to modern humans; it has a sheep element in that it argues that some animals still derive benefits from psi.

Sheep theories could be verified by locating psi alleles (gene forms). The search for specific alleles entails defining target phenotypes (observed variables) with sufficient precision that genotypes (corresponding genes) can be identified. Once candidate alleles are located, researchers can determine their specific actions, a process labeled “reverse phenotyping” (Schulze & McMahon, 2004). This allows more precise definitions of the phenotypes. If psi alleles were found, researchers could determine if these alleles had experienced positive or negative selection, testing the Freud/Palmer theory.

One method would be to gather DNA samples from families exhibiting high rates of anomalous experience. Researchers could then compare individual DNA from those reporting many experiences to those with no perceptions. Although one study found no linkage between paranormal belief and a dopaminergic gene (COMT), parallel theory-based studies could prove fruitful (Raz, Hines, Fossella, & Castro, 2008).

Selective breeding programs reveal associated phenotypes, a strategy that would help locate alleles. In a famous study, silver foxes were chosen for friendliness toward people and breed over 35 generations (Trut, 1999). Resulting tame foxes differed markedly from their wild relatives with regard to development, coat color, body features, and head shape. Breeding studies selecting for animal hypnotizability (tonic immobility) in rats and chickens demonstrated trait heritability and associated emotionality (Gordon, 1974).

If animal psi exists and has genetic basis, breeding studies selecting for psi performance should help locate psi alleles. Unfortunately, “taken as a whole, the evidence from animal psi research points to a psi effect that is small, not easily replicable, subject to fluctuations and decline effects, and whose source is largely unclear” (Dutton & Williams, 2009, p. 58). Even with these obstacles, success or failure to locate psi alleles has bearing on sheep theory; failure to find alleles reduces faith in the theory. An alternative study
design could have dog owners repeatedly breed pets thought to exhibit psi. Even if psi does not exist, this experiment could generate dogs with unusual phenotypes (perhaps parallel to popular conceptions of psi). Knowledge of these phenotypes could lead to alternate metaphors for exceptional human capacities.

Sheep hypotheses can be tested through multidimensional scaling analysis of community survey data. This method generates “maps” of variables with distance between variable clusters representing degree of correlation. Sheep theorists predict, and have found, that anomalous experiences are correlated with psychological well-being. Because psi’s benefits are thought to be greater than the evolutionary costs of psychosis, psi experiences are predicted to be more correlated with psychological well-being than with mental disorder (Sheep Hypothesis 1). Most sheep theorists believe that psi differs from psychopathology; multidimensional analysis should reveal psi experience clusters as separate from psychopathology clusters (Sheep Hypothesis 2).

An alternate, minority sheep position argues that psi is real but shares many alleles with mental disorder. Theorists note that schizotypy is defined in a way that includes psi experience. According to this argument, multidimensional scaling analysis should reveal psi experience variables within psychopathology clusters. This theory is weakened by the failure to locate alleles. If psi’s benefits are greater than psychopathology costs, modern gene-hunting methods should have located the beneficial alleles (Keller & Miller, 2006, but see Crispi, Summer, & Dorus, 2007).

Goat Theory

Skeptics argue that paranormal beliefs and experiences are rooted in human nature, yet products of cognitive dysfunction (Alcock, 1995; Alcock & Otis, 1980; Reed, 1988; for reviews see Irwin, 2009; Kelley, 2010). Although many goat theories are so vague that they cannot be tested, Schumaker (1990, 1995) provides a goat theory subject to empirical evaluation. He regards paranormal beliefs as equivalent to religious beliefs, since both involve delusional dissociation. He argues that paranormal/religious beliefs protect against excessive exposure to reality, and, as a result, confer psychological health. Anomalous experiences, such as ESP, tend to be pathological since they may deviate from socially accepted collective religious delusions (1990, p. 87). Schumaker (1990, p. vii) argues that paranormal belief, humankind’s strongest drive, “is taking us down the road to self-extinction.” As do many goat theorists, he believes that religion, although beneficial, has so many negative effects that it should be replaced by skeptical rationality (Dennett, 2006).

Goat theories tend to attribute anomalous experiences to specific cognitive dysfunctions. ESP, for example, indicates a deficiency of critical or cognitive ability (Alcock & Otis, 1980; Blackmore & Troscianko, 1985; Musch & Ehrenberg, 2002; see Irwin, 2009 for a critique). Out-of-body experiences result from malfunction of the mechanisms by which a person feels “in body” (Blackmore, 1982). Goat theories remove the need for paranormal explanations by hypothesizing that specific anomalous experiences are caused by corresponding physiological dysfunctions. OBEs, for example, are linked to schizophrenia due to abnormality of the angular gyrus (Powell, 2009, p. 114).

Goat hypotheses can be tested through multidimensional scaling analysis of community survey data. Each anomalous experience form should be most correlated with symptoms related to the psychopathological process producing it (Goat Hypothesis 1). Hypothesized relationships could have clinical applications since people reporting a particular anomalous experience might be at risk of the corresponding disorder. This argument allows two testable hypotheses regarding anomalous experience and pathology.

Sleep paralysis (involuntary immobility upon awakening) is often associated with hallucinations of demons, witches, or other spiritual forces. It “appears to be caused by a dysrhythmia of the sleep/wake cycle, controlled in part by adrenergic mechanisms” (Bell, Hildreth, Jenkins, & Garter, 1988, p. 289). Sleep paralysis may be governed by the same mechanisms producing anxiety, panic attacks, and hypertension (high blood pressure, HBP). African-Americans have a disproportionate incidence of sleep paralysis, HBP, stress, and panic attacks, suggesting shared alleles affecting anxiety. As a result, sleep paralysis may be predictive of HBP or anxiety disorders within a subset of African-Americans (Bell et al., 1988). Surveys asking about sleep paralysis might help identify of people at risk for HBP or anxiety disorder, allowing preventive treatment (Bennett, et al., 2004; James, 1994; James, Hartnett, & Kalsbeck, 1983). Studies contributing to
identifying at-risk individuals could be extremely fruitful since HBP contributes to reduced life expectancy among African-Americans (Wang & Wang, 2004). Sleep paralysis is hypothesized to be clinically predictive of, and highly correlated with, HBP and anxiety disorder symptoms (Goat Hypothesis 1a).

In parallel fashion, apparitions, ESP, OBEs, and NDEs have been found correlated with schizotypy. Goat theory argues that these correlations are due to psi’s equivalence to pathology. The theory hypothesizes that anomalous experiences are more correlated with schizotypy symptoms than with nonpathological predictive variables (Goat Hypothesis 1b). Identification of correlational variable clusters linked to psychosis (endophenotypes) could facilitate locating psychosis alleles (McClenon, 2011; Pearlson & Folley, 2008, p. 730). This line of research could contribute to the search for psychosis alleles and the development of better methods for identifying at-risk individuals.

According to Schumaker’s argument, psi experiences should be more correlated with pathological symptoms than with psychological well-being, since only accepted religions confer psychological health (Goat Hypothesis 2, opposing Sheep Hypothesis 1). Anomalous experiences should be more correlated with pathological symptoms than with religious experiences (Goat Hypothesis 3). This hypothesis coincides with studies indicating that belief in religious phenomena is generally not correlated with belief in paranormal phenomena (Rice, 2003). Religious experience should be more correlated with psychological well-being than with anomalous experience, thought equivalent to pathology (Goat Hypothesis 4). Multidimensional scaling should reveal anomalous experience clusters as within pathological symptom correlational clouds, since specific experiences are derived from specific pathologies (Goat Hypothesis 5, opposite of Sheep Hypothesis 2).

**Black Sheep Theories**

Black sheep theories use both sheep and goat arguments. These theories link anomalous experiences with both psychopathology and mental well-being and tend to be sympathetic to the field of parapsychology (Berenbaum, Kerns, & Raghavan, 2000). An emerging clinical paradigm portrays psychosis and spirituality as a continuum with some anomalous experiences linked to “healthy psychosis” (Clarke, 2010). Schizotypy, when associated with cognitive organization and adaptive religious beliefs, is associated with mental well-being (Goulding, 2005; McCreery & Claridge, 1995, 1996, 2002). Clinicians, using this approach, help clients determine “what is real and what is not” while avoiding the stigma of mental disorder (Clarke, 2010). Jackson (2010) notes that childhood trauma often results in psychopathology while benign psychosis is linked to positive childhoods. Anomalous experiences (suggesting benign psychosis) facilitate creative “paradigm shifting processes” contributing to psychological health (Jackson, 2010, Kennedy & Kanthamani, 1995).

The ritual healing theory allows testable black sheep hypotheses. This theory argues that anomalous experiences generate beliefs in spirits, souls, life after death, and magical abilities, the ideological foundations of shamanism (McClenon, 1997a, 2002a, 2011, 2012). Shamanic rituals provide evolutionary benefits due to hypnotic and placebo effects. As a result, shamanism, practiced over many millennia, selected for genotypes allowing modern forms of spirituality. Selected phenotypes (hypnotic capacity, absorption, dissociation, transliminality), allowing shamanism, became prevalent. Incidences of these phenotypes are correlated with anomalous experiences generating shamanic beliefs (McClenon, 1997a, 2002a, 2011). This theory portrays a continuum of experiences labeled as psychotic, schizotypal, anomalous, shamanic, and mystical—with increasing probability of benefit associated with shamanic and mystical perceptions (McClenon, 2011).

Folklore and anthropological and clinical evidence support this theory. Folklore studies indicate that universal forms of anomalous experience provide a basis for recurring folk religious beliefs (Hufford, 1982; McClenon, 1991, 1994, 1997a, 1997b, 2002a, 2002b, 2005). Paleolithic cave paintings, indicating shamanic altered states of consciousness, date back more than 30,000 years, sufficient time for shamanism to have evolutionary impacts (Lewis-Williams, 2002). During this time, the pace of evolutionary change has accelerated (Cochran & Harpending, 2009). Shamanic treatments involve hypnosis, a strategy proven effective for treating mild psychological disorders, childbirth complications, burns, blood loss following injury, and physiological processes beyond conscious control (Barber, 1984; Fromm & Nash,
A Community Survey of Anomalous Experiences

Such treatments would have evolutionary impact; hypnotizability, absorption, and religiosity are correlated and have genetic basis (Duke, 1969; Levin, Wickramasekera, & Hershberg, 1998; Morgan, 1973; Ott, Reiter, Henning, & Vaitl, 2005; Waller, Kojetin, Bouchard, Lykken, & Tellegen, 1990).

The ritual healing theory was evaluated through content analysis of 1,215 anomalous experience narratives. Findings supported theory hypotheses (McClenon, 2002b). The evidence also implied that hypothetical psi genes provide insufficient benefits to overcome their evolutionary costs.

Ritual healing hypotheses can be tested through multidimensional scaling analysis of community survey data. Many genes are regulated (switched on or off) by the effects of small carbon-based molecules governed by environmental factors. Alleles related to shamanism are hypothesized to be “switched on” by childhood stress/trauma and environmental conditioning, the same factors governing risk of psychosis. This argument coincides with findings presented by Rabeyron and Watt (2010). According to the ritual healing theory, anomalous experiences, absorption, related psychological variables (shamanic variables), and childhood difficulty should all be intercorrelated (Black Sheep Hypothesis 1). The ritual healing theory predicts that variables related to shamanism should be more correlated with schizotypy, anomalous experience, and childhood difficulty than are other variables regarded predictive (family income, life satisfaction, coping style, hostility; Black Sheep Hypothesis 2). Black sheep theory predicts that anomalous experience variables should be close, but not within, the psychological symptom cloud, as portrayed through multidimensional scaling. Like sheep theory, black sheep theory argues that psi is not equivalent to psychopathology (Black Sheep Hypothesis 3; equivalent to Sheep Hypothesis 2, opposite Goat Hypothesis 5).

Multidimensional Scaling Analysis of Psychological Symptoms

Mirowsky and Ross (1989) mapped 4,095 correlations of 81 psychological symptoms among 463 respondents in random sample surveys of El Paso, Texas, and Juarez, Mexico. Questionnaire items were regarded as well accepted survey measures of psychological distress. Multidimensional scaling analysis gives each variable a random location point and then adjusts each point in relation to its correlations with all other points. The points are shuffled until they best fit the correlations of each variable to every other variable. This method creates a kind of “map” of variables, showing clusters of variables with distance between variables representing degree of correlation.

Mirowsky and Ross (1989) used multidimensional scaling to evaluate hypotheses regarding relationships among psychological variables. A galaxy hypothesis suggested that psychological symptoms are clustered in distinct diagnostic categories. A nebula hypothesis predicted that symptoms are randomly distributed, not clustered by type. A spectrum hypothesis predicted that cluster types would overlap, shading into each other.

Study findings supported the spectrum hypothesis; diagnostic symptom groups were not discrete, but arrayed in a spectrum of overlapping clusters, arranged like a color wheel with ends connected (Mirowsky & Ross, 1989). Mirowsky (1990) argued that this pattern indicated a flaw in the Diagnostic and Statistical Manual (DSM) paradigm since the correlational map revealed no logical means for demarcating diagnostic categories. He suggested that DSM creators were like Renaissance witch hunters who made lists of signs defining witchcraft and achieved validity by demonstrating reliability among themselves. This argument coincides with the hypothesis that diagnostic categories represent reified concepts rather than valid diagnoses (Hyman, 2010) and that the search for schizophrenia alleles will be extremely frustrating, perhaps fruitless (Keller & Miller, 2006).

A fundamental question guided the present study design: How would multidimensional scaling portray the relationships among anomalous experience variables and Mirowsky and Ross’s (1989) psychological symptoms? Previous community surveys of anomalous experiences indicate that perception types are correlated with each other, suggesting a distinct, galaxy configuration (Palmer, 1979; Cardeña, Lynn, & Krippner, 2000). Sheep and black sheep theories predict a galaxy configuration (Sheep Hypothesis 2, Black Sheep Hypothesis 3) while goat theory predicts a spectrum pattern (Goat Hypothesis 5). A pilot community survey allows testing the two sheep hypotheses, five goat hypotheses, and three black sheep hypotheses using the SPSS ALSCAL program.
Questionnaires

A 193-item self-administered questionnaire was designed to evaluate the relationships between anomalous experiences, psychological symptoms, variables related to shamanism, and demographic variables and scales often correlated with mental disorder. The questionnaire measured: frequency of anomalous experience; psychological symptoms; Tellegen Absorption Scale; selected Transliminality, Boundary, and Dissociation Scale items; family income; blood pressure scale; Satisfaction With Life Scale; stress, childhood, and adolescence difficulty questions; John Henry Active Coping Scale; Hostility Questionnaire.

Frequency of anomalous experiences. Anomalous experience questions were derived from previous study findings. Between 1988 and 1997, Introduction to Anthropology undergraduate students at a predominately African-American college in North Carolina asked friends, relatives, and neighbors, “If you have had an extremely unusual experience, would you describe it?” The 1,446 respondents provided 1,578 anomalous experience accounts. These experiences were reliably classified into the following categories: apparitions (34.4%), paranormal dreams (9.6%), psychokinesis (PK; 9.0%), spiritual healings (6.1%), rootlore (experiences resulting from rituals derived, in part, from African occult traditions, McClenon, 2005; 5.0%), sleep paralysis (4.7%), waking ESP (4.3%), synchronistic events (3.2%), miscellaneous paranormal events (2.4%), occult events (associated with performances such as fortune telling or Ouija boards, 2.3%), out-of-body/near-death experiences (1.9%), and unidentified flying objects (1.7%). Other stories, deemed unusual by respondents but not regarded as anomalous, included folklore and religious experiences (6.9%), normal dreams (5.0%), and oral histories (3.7%; McClenon, 1994, 2002a, 2002b). Most rootlore and occult experience accounts could be classified within apparition, waking ESP, or PK categories.

Questionnaire items were designed to measure frequency of the 11 most common experiential forms. Respondents were asked, “How often have you had the following experience?” and provided the list: (a) apparition (perceiving something through sight, sound, or touch that you later found was not completely real), (b) waking ESP (perceiving something using a “sixth sense”), (c) paranormal dream (having a dream that you later found reflected actual events), (d) out-of-body experience (having the feeling that you were out of your body), (e) near-death experience (unusual perceptions during a time you thought you were close to death), (f) psychokinesis (perceiving that an object moved magically), (g) sleep paralysis (waking from sleep and finding that you could not move), (h) spiritual healing (being healed through a method not verified by modern medicine), (i) UFO (unidentified flying object in the sky), (j) other type of extremely unusual experience, (k) religious experience. Respondents were provided three frequency choices: “never,” “once or twice,” or “more often.”

Psychological symptoms. Mirowsky and Ross (1991) surveyed 91 psychological symptoms falling into seven diagnostic categories, many with subscales. Respondents were asked, “How often in the past 12 months have you” [and then provided a phrase specifying a symptom]. Response choices were “never,” “almost never,” “sometimes,” “fairly often,” and “often.” The present study omitted questions requiring a professional interviewer. An additional OCD item (OCD3) was devised and added, based on the wording on page 462 of the DSM IV-TR (American Psychiatric Association, 2000). As a result, the study questionnaire included 88 self-administered symptom questions associated with 11 diagnostic categories (see Appendix).

Among psychological symptoms, 12 items measure the diagnostic category schizophrenia. Ten items (Schizo1 to Schizo10) ask about schizotypy experiences and 2 items (Schizo11 and Schizo12) pertain to cognitive disorganization. This diagnostic category is labeled schizotypy in the present report, even though the scale does not contain items asking about introverted anhedonia or impulsive nonconformity, variables typically included on schizotypy scales (Claridge & Beech, 1995; Chapman, Chapman, & Kwapiel, 1995).

Tellegen Absorption Scale (Absorp). The 34-item Tellegen Absorption Scale (TAS) measures openness to absorbing and self-altering experiences (Tellegen & Atkinson, 1974; Cronbach α = .84).
Although measurement of absorption is influenced by context, many studies have found TAS to be correlated with hypnotic suggestibility (Heap, Brown, & Oakley, 2004).

**Transliminality (Translim), Boundary (Boundary), and Dissociation (Dissoc) scales.** The 29-item Transliminality Scale was designed to measure tendency for psychological material to cross thresholds into or out of consciousness (Lange, Thalbourne, Houran, & Storm, 2000; Thalbourne & Delin, 1994; \( \alpha = .87 \)). The 146-item Boundary Questionnaire was designed to quantify cognitive boundaries; people with thin boundaries are thought to slide easily from one feeling to another and to be more vulnerable and open to new ideas (Hartmann, 1991; \( \alpha = .84 \)). The 28-item Dissociative Experience Scale was designed to quantify propensity for dissociative experiences (Carlson, et al., 1993; Bernstein & Putnam, 1986; \( \alpha = .92 \)).

In order to select questionnaire items reflecting hypnotic suggestibility, the Harvard Group Scale of Hypnotic Susceptibility, Form A (Shor & Orne, 1962) was administered to 154 undergraduate students. Transliminality, Boundary, and Dissociation scales were administered to these students on alternate occasions. The 15 questionnaire items most correlated with Harvard Group Scale scores were selected for use on the study questionnaire (4 transliminality items, 4 dissociation items, 7 boundary items; see Appendix).

**Income and blood pressure questions (Income and HBP).** The study used Centers for Disease Control (2008 and previous years) survey questions regarding annual household income (ordinal scale). Annual household income and socioeconomic status, at birth, are predictive of psychosis incidence (Werner, Malaspina, & Rabinowitz, 2007). The sum of three CDC standardized ordinal-scale questions regarding high blood pressure (HBP) generated the HBP scale.

**Satisfaction With Life Scale (Satisfac).** The 5-item Satisfaction With Life Scale measures global life satisfaction. It indicates psychological well-being and has been found to be strongly negatively correlated with clinical measures of distress (Pavot & Diener, 1993; mean \( \alpha = .78 \)). See Appendix for questionnaire items.

**Stress, childhood, and adolescence difficulty questions.** Respondents were asked to evaluate the sentence: “I feel that my life is more stressful than that of most people” (Stress). Two items, taken from Hartmann’s (1991) Boundary Questionnaire, were also included: “I feel that I had a difficult and complicated childhood” (DifChild) and “I feel I had a difficult and complicated adolescence (DifAdol).” Response choices were “never,” “rarely,” “generally,” and “often.”

**John Henry Active Coping Scale (JHS).** The John Henry Active Coping Scale consists of 12 statements evaluated by the respondent from 1 to 5 as “completely false” to “completely true” (Ford, Hill, Butler, & Havstad, 2002; \( \alpha \) varied from .66 to .80). This scale predicts propensity for high blood pressure in some populations, particularly lower socioeconomic status (SES) African-Americans, and is correlated with a wide array of health status outcomes (Bennett et al., 2004; James, 1994; James, Hartnett, & Kalsbeck, 1983).

**Hostility Questionnaire (Anger).** The Hostility Questionnaire is a 46-item scale subdivided into cynicism, anger, and aggression subscales, predictive of both psychological health and physical mortality (Williams & Williams, 1994, pp. 5–11; \( \alpha = .66 \)). Twelve items, regarded as indicative of mental health and found in pilot studies to be correlated with high blood pressure (HBP), were selected for use on the study questionnaire.

**Questionnaire Administration**

As part of class projects, anthropology students at a predominately African-American college in North Carolina were trained to administer an approved Institutional Review Board (IRB) consent form and the study questionnaire. They recruited friends, family members, and neighbors to participate in this study. After signing the consent form, participants privately completed the questionnaire, sealed it in a provided envelope, and returned it to the student. The student then interviewed the respondent regarding unusual and stressful experiences using questions designed to evaluate validity of anomalous experience items. The student transcribed the interview and assigned the data an anonymous code number.
Sample Characteristics

Respondent demographics reflect the composition of students administering the questionnaire during the study period (2001–2006). Of 965 respondents providing information, 73% were African-American and 69% were female. Mean age was 33 years. These demographics correspond with those of the student interviewers and the predominantly African-American community surrounding the college campus.

Estimated median annual family income was $27,372. On average, African-Americans reported less annual family income (estimated median = $23,022) than did Caucasians (estimated median = $41,557; \( \chi^2 (4, N = 890)= 56.1, p < .001; \) African-Americans earned 66% of Caucasian income). During the study period, median annual family income in the USA varied from $49,455 to $50,255 (2007 dollars), with African-American families earning, on average, 62% of non-Hispanic White family earnings. These values imply that the sample median family income was in the bottom quartile of the American population. Among those providing information, 13.5% were students and 40.1% were unemployed.

The sample was drawn from a predominantly African-American area of North Carolina with high poverty rate (over 20%), little economic growth, high unemployment, and problems with mental health care services. Sociological status is inversely correlated with mental health. In general, African-Americans are about 3 times more likely to be diagnosed with schizophrenia than are White Americans, a tendency only partially reduced by controlling for socioeconomic status at birth (Bresnahan et. al., 2007).

Relatively high percentages of respondents reported specific anomalous experiences “once or twice” or “more often”: apparitions (40%), paranormal dreams (66%), waking ESP (21%), OBE (29%), NDE (14%), PK (34%), sleep paralysis (35%), spiritual healing (35%), UFO (10%), other (22%), religious experience (23%). These frequencies coincide, in general, with those reported by Greeley (1987) in a national USA sample and McClenon (1994, 2002a) in random samples of three USA colleges. Rates of anomalous experience reported by students at the predominantly African-American college were equivalent to those reported at two predominantly Caucasian colleges (except for sleep paralysis, which is reported more frequently among African-Americans).

Differences regarding gender and ethnicity produced 12 statistically significant t values (\( p < .01 \)). African-Americans, compared to Caucasians, reported greater incidence of sleep paralysis, \( t(905) = 3.5, p < .001 \); spiritual healing, \( t(894) = 3.6, p < .001 \); schizotypy, \( t(858) = 2.7, p < .007 \); paranoia, \( t(861) = 5.9, p < .001 \); lower incidence of UFO experiences, \( t(899) = 3.9, p < .001 \); absorption, \( t(776) = 2.7, p < .007 \); and difficult adolescence, \( t(896) = 5.7, p < .001 \). Females reported higher rates of depression-mood, \( t(796) = 2.7, p < .006 \); anxiety-mood, \( t(857) = 3.1, p < .002 \); and phobia, \( t(882) = 2.6, p < .009 \); but fewer alcohol problems \( t(875) = 3.6, p < .001 \), and lower family income, \( t(766) = 4.5, p < .001 \).

Family income was negatively, but slightly, correlated with schizotypy symptoms, \( r(903) = -.16, p < .001 \). Controlling for income rendered the schizotypy/ethnicity relationship nonsignificant (both chi square and multiple regression analysis; \( p > .05 \)). In general, the literature indicates that socioeconomic status is inversely correlated with mental disorder (Hudson, 2005).

Multidimensional Scaling

Analysis of Hypotheses

Figure 1 portrays the diagnostic categories as a spectrum of overlapping clusters, arranged like a “color wheel” with ends connected—replicating Mirowsky and Ross (1989). This configuration supports their spectrum hypothesis (McClenon, 2012).

Schizotypy symptoms are scattered in the upper-right quadrant of this wheel with anomalous experience variables arrayed above them. Psi-related experiences (apparitions, ESP, NDE, “Other” anomalous experiences, PK) are close to schizotypy symptom variables but not within the symptom cloud, a galaxy pattern (supporting Sheep Hypothesis 2, Goat Hypothesis 1b, Black Sheep Hypothesis 3; failing to
A Community Survey of Anomalous Experiences

Support Goat Hypothesis 5). Sleep paralysis (Sleeppar) is not highly correlated with symptoms indicating anxiety disorder (anxiety-mood, AM1 to AM5; anxiety-malaise, Amal1 to Amal8; panic, panic1 to panic5. This fails to support Goat Hypothesis 1a.

**Derived Stimulus Configuration**

**Euclidean distance model**

*Figure 1. Multidimensional analysis of anomalous experiences and psychological symptoms.*

Although the anomalous experience variables (apparitions, waking ESP, paranormal dreams, OBE, NDE, PK, sleep paralysis, spiritual healing, UFOs, and “Other”) do not form a particularly tight cluster, a scale made up of these variables has a Cronbach’s $\alpha$ of .80. Cronbach’s $\alpha$ indicates a scale’s internal consistency or reliability; in the social sciences, scores above .7 are regarded as “acceptable” and .8 to .9 are labeled “good” (George & Mallery, 2003). The anomalous experience scale is acceptable, almost good (McClenon, 2012).

This pattern differs from what some sheep theorists might expect. Folklore attributes the anomalous experience forms to widely different processes. Apparitions are thought to originate with spirits of the deceased. OBEs reflect the soul’s ability to leave the body. ESP implies information gained paranormally. UFOs are associated with extraterrestrial visitors. The high degree that these variables are intercorrelated suggests that processes other than those specified by folklore govern their incidence.

*Figure 2 portrays correlations among absorption, dissociation, transliminality, and boundary variables, childhood and adolescent difficulty, present stress, and anomalous experiences. Childhood and adolescent difficulty are highly correlated with each other and with many of the absorption and dissociation variables (forming a spectrum cloud with regard to these variables). Absorption, dissociation, boundary, and transliminality variables form a shamanic variable cloud, distinct from the loose anomalous experience cloud (galaxy pattern). Shamanic variables are correlated most highly with paranormal dreams, apparitions,
ESP, OBE, and PK. The configuration fits the ritual healing theory, which suggests that childhood difficulty switches on alleles governing shamanic variables which, in turn, trigger anomalous experiences. The configuration supports Black Sheep Hypotheses 1.

Figure 2. Anomalous experience variables and psychological variables.

Figure 3 portrays relationships between diagnostic categories, anomalous experience variables and scales regarded as related to mental health. The John Henry Scale ($\alpha = .883$) is in the upper left quadrant; Satisfaction Scale ($\alpha = .87$) in the lower left quadrant; Anger Scale ($\alpha = -.10$) is in the lower left and right quadrants. These scales are arranged in clusters to the left of the mental disorder diagnostic cluster (generally upper right quadrant). The Anger Scale failed to achieve an acceptable level of internal consistency, revealing a very loose cluster. The anomalous experience variables are close to or within the diagnostic and shamanic variable clusters, made up of Dissociation, Boundary, Transliminality, and Absorption Scales (supporting Goat Hypothesis 2).

As predicted by Black Sheep Hypotheses 1, Pearson correlations were significant ($p < .001$, $N$ varying from 784 to 933) between schizotypy (Schizo) and App (.36), ESP (.32), Pdream (.31), OBE (.39), NDE (.36), PK (.39), Sleeppar (.27), Other (.40), Healing (.32), UFO (.18), Relig (.28), DifChild (.28); as well as DifAdol (.30), Stress (.16), Absorption (.54), Boundary (.55), Translim (.51), and Dissoc (.50).

The figure suggests that, as specified by the ritual healing theory, diagnostic categories, psychological variables, and anomalous experience variables are affected by childhood and adolescent difficulty. Other variables (John Henry Scale, Life Satisfaction, Anger) are not as closely correlated with psychological symptoms or anomalous experience. Anomalous experience variables are more correlated with diagnostic categories than with scales thought to be correlated, Life Satisfaction (Sat1–Sat5) and the Anger Scale.
A Community Survey of Anomalous Experiences

(Albert1–Albert12). This pattern refutes Sheep Hypothesis 1, supports Goat Hypothesis 2, and supports Black Sheep Hypothesis 2.

\[ \text{Dimension 1} \]
\[ \text{Dimension 2} \]

**Derived Stimulus Configuration**

**Euclidean distance model**

![Diagram showing multidimensional scaling of diagnostic variables, variables related to shamanism, anomalous experience variables, John Henry Scale (JHS), Hostility Scale (Anger), and Satisfaction With Life Scale (Sat).](image)

Figure 3. Multidimensional scaling of diagnostic variables, variables related to shamanism, anomalous experience variables, John Henry Scale (JHS), Hostility Scale (Anger), and Satisfaction With Life Scale (Sat).

Religious experiences (Relig) are not more correlated with Life Satisfaction (Sat1-Sat5) than are other anomalous experiences, refuting Goat Hypothesis 4. Sleep paralysis (Sleeppar) is not as correlated with panic or anxiety disorders as are other variables, refuting Goat Hypothesis 1a.

Figure 4 includes variables found predictive of mental disorder in the literature: income, ethnic ID (Ethnicid, coded as African-American or Caucasian/White), gender, age, high blood pressure (HBP), satisfaction with life (Satisfac), John Henry Scale (JHS), and Anger Scale (Anger). Also shown are diagnostic scales and anomalous experience variables. The figure portrays a variable cloud that includes alcohol, paranoia, difficult adolescence, difficult childhood, anxiety-mood, phobia, panic disorder, depression-mood, OCD, anxiety-mal (anxiety-malaise), schizotypy, manic behavior (lower right quadrant with Schizo in the upper right quadrant). The anomalous experience variables form a very loose cloud about the diagnosis cloud (upper right quadrant with UFO and PK in the upper left quadrant). The paranormal experiences variables are less correlated with satisfaction with life then are the diagnostic variables (refuting Sheep Hypothesis 1 and supporting Goat Hypothesis 2).

Sleep paralysis (Sleeppar) was not highly correlated with high blood pressure (HBP) and variables such as Panic or Phobia (refuting Goat Hypothesis 1a). As noted previously, religious experience (Relig)
is not more correlated with the Satisfaction With Life Scale (Satisfac) than are the anomalous experience variables (refuting Goat Hypothesis 4).

**Derived Stimulus Configuration**

**Euclidean distance model**

*Figure 4.* Multidimensional analysis of diagnostic variables, childhood/adolescent difficulty, anomalous experience variables, and other scales.

Figure 5 portrays multidimensional scaling for schizotypal experiences, shamanic variables, anomalous experience variables, stress, and childhood/adolescent difficulty. The figure reveals a galaxy pattern with schizotypy as a loose cloud (schizo11 and schizo12 are outside of the cloud); shamanic variables form another cloud (upper right quadrant, but extending into the upper left and lower right quadrants). The anomalous experience variables form a very loose cloud or perhaps two clouds (lower left quadrant with App, ESP, Pdream in the lower right quadrant). This configuration reveals schizotypy, anomalous experience, and shamanic variables as distinct entities (supporting Sheep Hypothesis 2 and Black Sheep Hypothesis 3, refuting Goat Hypothesis 5). PK, very close to the schizotypy cloud, is an exception.

Although some schizotypy symptoms refer directly to anomalous experiences (Schizo2—auditory hallucinations; Schizo3—visual hallucinations; Schizo3—visions), the Schizo12 symptom “had useless thoughts that kept running through your mind” was most correlated with apparitions (App), ESP, and paranormal dreams (Pdream). This study portrays unexpected correlations between the disorganized aspects of schizotypy and anomalous experience.

Regression analysis reveals scales predictive of schizotypy. A BTD scale was created by summing Boundary, Transliminality, and Dissociation items ($\alpha = .81$). An anomalous experience scale was created by summing anomalous experiences ($\alpha = .80$). Schizotypy was highly correlated with BTD ($R = .59$, $N$ =
A Community Survey of Anomalous Experiences

814) and anomalous experience ($R = .60, N = 864$). With schizotypal experience as the dependent variable, the BTD Scale and Unusual Experience Scale explained almost half of schizotypal variance ($R = .69, R^2 = .47, \text{Adj. } R^2 = .47, \text{S.E.} = 4.71$). People with difficult childhoods, anomalous experiences, and shamanic propensities are likely to report schizotypy symptoms.

**Figure 5.** Schizotypal experience, psychological variables, anomalous experience variables, stress variables.

**Discussion**

This study replicated Mirowsky and Ross’s (1989) findings that mental disorder symptoms do not cluster in ways that allow clear demarcations (McClenon, 2012). This finding coincides with criticisms of the DSM paradigm and suggests that diagnostic categories, such as schizophrenia, do not reflect underlying genetic structures (Adriaens, 2008).

Study findings allow evaluation of sheep, goat, and black-sheep hypotheses. Sheep Hypothesis 1 and Goat Hypotheses 4 and 5 were not supported. Sheep Hypothesis 2 and Goat Hypotheses 2 and 3 were supported. Goat Hypothesis 1b regarding schizotypy-anomalous experience correlations was supported but Goat Hypothesis 1a regarding sleep paralysis was not supported. All three black sheep hypotheses were supported. In summary, 50% of sheep hypotheses, 50% of goat hypotheses, and 100% of black sheep hypotheses were supported.

Black sheep findings coincide with the idea that spirituality works as a coping strategy (Pargament, 1997). Spirituality ameliorates the effects of stress, regardless of religious affiliation, with benefits derived from reducing negative affect (Kim & Seidlitz, 2002). Spirituality and core spiritual experiences have a
buffering quality, powerful enough to reduce medical symptoms (Kass, Friedman, Leserman, Zuttermeister, & Benson, 1991).

These findings support an anthropological paradigm suggesting that shamanic ritual can reduce psychosis incidence (McClenon, 2002a; Winkelman, 2010). Although religion is connected with psychosis, shamanism has characteristics that may prevent psychosis. Shamanic traditions describe training programs that help at-risk people increase their cognitive organization. People reporting difficult childhoods, anomalous experiences, and shamanic characteristics could benefit from ritual training designed to increase cognitive integration.

Study Hypotheses and Genetics Research

The study plan was devised in 2000 with the hope of contributing to genetics research. Multidimensional scaling of community survey data can reveal variable clusters correlated with psychosis, suggesting associated phenotypes with underlying alleles. With adequate definition of phenotypes, genetics researchers can locate these alleles (Lenzenweger, 2010).

Unfortunately, progress toward finding psychosis alleles has been disappointing (Wade, 2009). Keller and Miller (2006) provide a theory explaining this outcome. They suggest that, rather than derived from specific alleles, many mental disorders originate with random mutations. They argue that cognition has “added on” characteristics since modules developing during early eras of human evolution affect modules evolving during later stages. According to their theory, mutations disrupting “upstream” modules can result in a cascade of failures, impacting “downstream” modules. The diagnostic symptoms of psychosis, for example, may reflect cascade failures within the upstream/downstream cognition structure rather than specific psychosis alleles. Although harmful mutations are eventually eliminated by evolutionary costs, new mutations continually trigger similar cascade failures due to the physiology of consciousness. This theory explains “the data on mental disorder prevalence rates, fitness costs, the likely rarity of susceptibility alleles, the increased risks of mental disorders with brain trauma, inbreeding, and paternal age” (Keller & Miller, 2006, p. 385).

Keller and Miller’s (2006) theory explains balance theory failures. If sheep theory is valid, modern gene hunting methods should have found underlying psi/psychosis alleles. If Schumaker’s goat theory is correct, psi incidence should have declined over time. Neither theory’s predictions have been supported.

Black sheep theory combines sheep and goat arguments. It suggests that psi is real but spontaneous incidence is correlated with psychological symptoms, childhood difficulty, and shamanic variables. This theory suggests that, even if few psychosis alleles exist, shamanic alleles can be located. A gene influencing absorption has already been found (Ott, Reiter, Henning, & Vaitl, 2005).

Black sheep theories acknowledge religion’s benefits as well as dysfunctions. Although psychosis/spirituality has evolutionary costs, spirituality and associated anomalous experiences generate folk beliefs and coping skills. These experiences provide a basis for shamanism.

Shamanism involves controlling anomalous cognitive skills. Humans vary in their control over these capacities. Although most people have no control over spontaneous perceptions, some elicit hypnotic, out-of-body experiences. This capacity, coupled with trance performance, allows shamanism.

Group experiences shape folk beliefs. Haunting, poltergeist, and possession experiences can be influenced by ritual. Small groups (sitter, prayer, spiritualist, cults, sects) experience collective anomalous events, affected by belief. Shamans achieve sufficient cognitive discipline that they can provide trance performances, generating placebo and hypnotic benefits.

Shamanism and its modern variations create innovative folk beliefs regarding group problems. Black sheep theory argues that psi’s power is derived from associated beliefs rather than direct evolutionary benefits or contribution to scientific progress.

Although religious and scientific elites feel threatened by innovation, shamanic alleles insure that future people will find ways of thinking that fit their needs. Rather than leading to humanity’s extinction (Schumaker, 1990, p. 87), the black sheep theory predicts that these alleles will contribute to therapeutic processes. As noted by Plato, the forms of madness labeled spirituality “will prevail with the wise, though not with the learned” (Harvey, 1996, p. 120).
A Community Survey of Anomalous Experiences

Study Limitations

Exploratory methodology allows many possibilities for bias. These include self-reporting of psychological symptoms and variables, selection of participants by student interviewers, use of single-item measures for various variables, and ordinal levels of measurement of most variables. The degree that present findings can be generalized is unclear. The study population differs markedly from random national USA samples.

Multidimensional scaling does not identify causal or spurious relationships. Large sample size limits the applicability of tests of statistical significance. Ordinal level data limits usefulness of multiple regression analysis. As with any theory-based study, a variety of alternate explanations exist for correlative findings. Verifications of hypotheses do not prove a theory’s validity but provide guidance for future studies.

Correlations among variables might be attributed, in part, to parallel response biases and the tendency for some people to choose extreme responses. Alternate ordering of questionnaire items might affect correlative results. Although previous researchers acknowledge similar methodological problems, they report recurring patterns within their data, similar to those found in the present study (Mirowsky & Ross (1989; Palmer, 1979; Rabeyon & Watt, 2010). These patterns increase our faith in methodological validity.

Correlations among psychological variables may differ among cultures. Correlations between anxiety and depression, depression and irritability, and anxiety and irritability varied widely in data from Denmark, India, Colombia, Nigeria, England, Russia, Czechoslovakia, Taiwan, and USA and between New York African-Americans and New York Caucasians (Leff, 1981, pp. 47–49). Studies seeking to define psychological phenotypes require replication among multiple ethnic groups.

Conclusions

Incidence of psychological symptoms, variables related to shamanism, propensity for anomalous experience, and childhood/adolescent difficulty were significantly correlated. Correlational patterns could be explained as the result of childhood difficulty triggering alleles related to both psychopathologies and shamanic variables. Although alternative explanations exist, findings are in harmony with the idea that shamanism, practiced for many millennia, shaped phenotypes related to hypnotic and placebo treatment.

If studies of other ethnic groups replicate these findings, this line of research could have practical applications. Schizotypal individuals reporting frequent anomalous experiences and revealing shamanic phenotypes could benefit from programs designed to increase their cognitive self-control through spiritual practice.

Studies of this type can shed light on the spirituality/psychosis relationship. Anomalous experiences seemingly contribute to genetically-based systems that alleviate the effects of stress and trauma.

References


Virginia Beach Psychiatric Center
1100 First Colonial Road
Virginia Beach, VA 23454
beinghere@hotmail.com

Abstracts in Other Languages

French

UNE ENQUÊTE COMMUNAUTAIRE SUR LES EXPERIENCES ANOMALES : ANALYSE DE CORRELATION DES HYPOTHESES EVOLUTIONNISTES

RESUME: Un questionnaire, administré à un échantillon à prédominance afro-américaine dans le Nord-Est de la Caroline du Nord (N = 965), a sondé l’incidence des expériences anormales, des symptômes psychologiques, des variables psychologiques liées au chamanisme, au bien-être psychologique et à des variables démographiques. Une analyse d’échelle multidimensionnelle a permis une évaluation des hypothèses dérivées des théories du mouton, de la chèvre et du « mouton noir ». Les théories du mouton prédisent que les expériences paranormales fournissent des avantages directs en termes de survie, grâce au psi. Les théories de la chèvre affirment que le psi n’existe pas ; les expériences anormales sont associées avec la psychopathologie et ne fournissent aucun bénéfice direct pour l’évolution. Une théorie du « mouton noir » possède des éléments des moutons et des chèvres ; elle décrit un continuum de la psychose à la spiritualité, avec des bénéfices dérivés de la spiritualité ; le psi pourrait exister mais ne pas fournir des bénéfices évolutifs directs suffisant pour surmonter les coûts psychopathologiques. Au sein du paradigme du mouton noir, la théorie de la guérison rituelle affirme que des variables chamaniques et les génotypes associés facilitent les aptitudes à faire face et les effets hypnotiques / placebo. Les résultats de cette étude fournissent un soutien mitigé pour les hypothèses du mouton et de la chèvre mais un soutien total aux hypothèses du mouton noir.

Spanish

UNA ENCUESTA COMUNITARIA SOBRE EXPERIENCIAS ANÓMALAS: UN ANÁLISIS CORRELACIONAL DE HIPÓTESIS EVOLUTIVAS

RESUMEN: Un cuestionario, administrado a una muestra predominantemente afro-americana en el noreste de Carolina del Norte (N = 965) investigó la incidencia de experiencias anómalas, síntomas psicológicos, variables psicológicas relacionadas con el chamanismo, escalas relacionadas con el bienestar psicológico, y variables demográficas. Un análisis dimensional de escalas permitió evaluar hipótesis extraídas de teorías de ovejas, cabras, y “ovejas negras”. La teoría de ovejas predice que las experiencias paranormales proporcionan ventajas de supervivencia directas, derivadas de psi. La teoría de cabras argumenta que psi no existe; las experiencias anómalas están asociadas con psicopatología y no proporcionan ningún beneficio evolutivo directo. La teoría “ovejas negras” tiene elementos ovinos y caprinos, postulando un continua psicosis-espiritualidad, con beneficios derivados de la espiritualidad; es posible que psi exista pero no proporciona beneficios evolutivos directos suficientes para superar los costos psicopatológicos. Dentro del paradigma ovejas negras, la teoría de la curación ritual postula que las variables
chamánicas y sus genotipos asociados facilitan habilidades de afrontamiento y efectos hipnóticos/placebo. Los resultados del estudio proporcionan apoyo mixto para las hipótesis ovina y caprina, pero son totalmente compatibles con la hipótesis de ovejas negras.

German

EINE GEMEINDE-UMFRAGE ZU ANOMALEN ERFAHRUNGEN:
KORRELATIONSA NALYSE EVOLUTIONÄRER HYPOTHESEN


APPENDIX

Questionnaire Items and Scales

Psychiatric Symptoms Categorized by Psychiatric Diagnosis

Schizophrenia/Schizotypy (Schizo)
S1: Felt that your mind was dominated by forces beyond your control
S2: Heard voices without knowing where they came from
S3: Seen things or animals or people around you that others did not see
S4: Had visions or seen things other people say they cannot see
S5: Felt that you were possessed by a spirit or devil
S6: Felt you had special powers
S7: Felt that you did not exist at all, that you were dead, dissolved
S8: Seemed to hear your thoughts spoken aloud - almost as if someone standing nearby could hear them
S9: Felt that your unspoken thoughts were being broadcast or transmitted, so that everyone knew what you were thinking
S10: Felt afraid that you might do something seriously wrong against your own will
S11: Had unusual thoughts that kept bothering you
S12: Had useless thoughts that kept running through your mind

Paranoia (Paranoia)
P1: Felt it was safer to trust no one
P2: Believed you were being plotted against
P3: Been sure that everyone was against you
P4: Felt that people were saying all kinds of things about you behind your back
P5: Felt you had enemies who really wished to do you harm
P6: Been very suspicious, didn’t trust anybody
Depression-Mood (DepMood)
DM1: Thought about taking your own life
DM2: Wondered if anything was worthwhile anymore
DM3: Felt that nothing turned out for you the way you wanted it to
DM4: Felt that you deserved to be punished
DM5: Felt that others would be better off if you were dead
DM6: Felt that you have done something evil or wrong
DM7: Wished you were dead
DM8: Felt very bad or worthless
DM9: Blamed yourself for something that went wrong
DM10: Felt completely hopeless about everything
DM11: Felt lonely
DM12: Felt like crying
DM13: Felt guilty about things you did or did not do
DM14: Felt useless
DM15: Lost your temper
DM16: Been in low spirits
DM17: Brooded over unpleasant thoughts or feelings
DM18: Just didn’t care what happened to you
DM19: Been moody and unhappy
DM20: Felt completely helpless

Manic (Manic)
Manic1: Had times when exciting new ideas and schemes occurred to you one after another
Manic2: Became so excited that your thoughts raced ahead faster than you could speak them
Manic3: Felt so great (excited, talkative or active) that it was difficult to concentrate

Depression-malaise (DepMal)
DMal1: Became very quiet and didn’t talk to anyone
DMal3: Shown no interest in anything or anybody
DMal4: Had trouble concentrating or keeping your mind on what you were doing
DMal5: Kept losing your train of thought
DMal6: Felt that your mind did not work as well as it used to
DMal7: Had periods of feeling blue or depressed that interfered with your daily activity
DMal8: Had periods of days or weeks when you couldn’t take care of things because you couldn’t “get going”
DMal9: Felt confused: had troubled thinking
DMal10: Got angry and afterward felt uncomfortable, like getting headaches, stomach pains, cold sweats and things like that
DMal11: Began having trouble remembering things
DMal12: Had trouble staying asleep
DMal13: Had trouble with waking up too early and not being able to fall asleep again
DMal14: Had trouble with oversleeping: that is, sleeping past the time you wanted to get up
DMal15: Troubled by feeling tired all the time
DMal16: Been bothered by nightmares
DMal17: Had poor appetite
DMal18: Felt weak all over
DMal19: Experienced any weight loss of 10 lbs. (5 kg) or more over the past year, without going on special diets

Anxiety-mood (AnxMood)
AM1: Worried a lot about little things
AM2: Felt anxious about something or someone
AM3: Were easily irritated
AM4: Been bothered by being irritable, fidgety, or tense
AM5: I am a person who is the worrying type

**Panic (Panic)**
Panic1: Felt afraid to leave the house because you were afraid something might happen to it
Panic2: Been afraid to be in closed places
Panic3: Feared something terrible would happen to you
Panic4: Had special fears that kept bothering you
Panic5: Feared being robbed, attacked, or physically injured

**Anxiety-malaise (AnxMal)**
AMal1: Had trouble with your muscles twitching or jumping
AMal2: Had trouble falling asleep
AMal3: Had cold sweats
AMal4: Had dizziness
AMal5: Had shortness of breath when you were not exercising or working hard
AMal6: Had your hands tremble
AMal7: Had your heart beating hard when you were not exercising or working hard
AMal8: Suddenly feel hot all over
AM1 (Anxiety-behavior): Had periods of such great restlessness that you could not sit in a chair for very long

**Anxiety-Obsessive (OCD)**
OCD1: Had to repeat an act over and over again though it is hard to explain to others why you did it
OCD2: Found yourself doing the same things over and over again to be sure they were right
OCD3: Spent more than an hour day focusing on persistent anxieties that do not involve real life problems but are products of your mind

**Alcoholism (Alcohol)**
Alcohol1: Missed work or been late to work because of drinking
Alcohol2: Had arguments with your family because of your drinking
Alcohol3: Had trouble with your health because of drinking

**Phobia (Phobia)**
Phobia 1: Had a sudden feeling, not based on a real cause, of intense apprehension, fearfulness or terror
Phobia 2: Felt you must avoid certain places or situations from which escape might be difficult or embarrassing
Phobia 3: Had an unreasonable fear, not associated with a real cause, regarding snakes, dogs, insects, or other animals
Phobia 4: Had an unreasonable fear, not associated with a real cause, regarding heights, storms, lightning, or water

**Selected Transliminality Scale Items (Translim)**
T1. At time I perform certain little rituals to ward off negative influences.
T2. I have experienced an altered state of consciousness in which I felt that I became cosmically enlightened.
T3. I have experienced an altered state of consciousness which I believe utterly transformed (in a positive manner) the way I looked at myself.
T4. I sometimes have a feeling of gaining or losing energy when certain people look at me or touch me.
Selected Dissociative Experience Scale Items (Dissoc)
D1. I find that sometimes I am listening to someone talk and I suddenly realize that I did not hear part or all of what was said.
D2. I have become so involved in a fantasy or daydream that it feels as though it were really happening to me.
D3. I have found that I am able to ignore pain.
D4. I find that I sometimes sit staring off into space, thinking of nothing, and am not aware of the passage of time.

Selected Boundary Questionnaire Items (Boundary)
B1. I can visualize something so vividly that it is just as though it is happening right in front of me.
B2. I have dreams and daydreams or nightmares in which I see isolated body parts – arms, legs, heads, and so on.
B3. In my dreams, people sometimes merge into each other or become other people.
B4. I believe I am influenced by forces which no one can understand.
B5. I think I would be a good fortune teller or a medium.
B6. When I recall a conversation or a piece of music, I hear it just as though it were happening there again right in front of me.
B7. I can see fields of energy around people.

Satisfaction With Life Scale (Satifac)
1. In most ways my life is close to my ideal
2. The conditions of my life are excellent.
3. I am satisfied with my life.
4. So far I have gotten the important things I want in life,
5. If I could live my life over, I would change almost nothing.
ABSTRACT: Data on telepathic group communication of emotions, as evoked by slide pictures, were analysed with a view to identifying critical picture characteristics. Performance was related to 6 psychological picture scales as well as to 2 scales measuring the pictures' tendency to evoke electrodermal (EDA) and heart-rate (HR) responses, respectively. All 8 picture scales were merged into a composite scale, reflecting negative arousal (NA). Based on this scale, a negative arousal discrimination (NAD) scale was constructed. An old data set, obtained from 845 participants, and a new one, obtained from 652 participants, were analysed, both separately and together, with relative hit rate (hit rate in relation to response tendency) as the performance measure. Significant interstudy reliability was established. Participants discriminated between 2 types of negative pictures: those with high and those with low negative arousal potential. For negative pictures, the NAD scale was significantly positively skewed, with a remarkably small p value for the total data set (p = .000001). This skewness could largely—but not fully—be accounted for in terms of a significant (p = .0003) negative correlation between the NAD scores and number of receivers. The major findings tended to be clearer in the new data set than in the old one. Various “natural” explanations of the positive results are discussed and dismissed as unlikely. A simplified experiment for future replication experiments is suggested.

Keywords: group telepathy, emotion, arousal, slide pictures, group size

Despite their practical advantage of being able to provide a large amount of data in a short period of time, group experiments on extrasensory perception (ESP) are relatively rare. One reason is probably the common opinion that group testing is, for some reason, inferior to individual testing in producing positive results (see, e.g., Rhine, 1947/1971, p. 40). However, there is little empirical support for this view. Thus, not only negative results (e.g., Haight, Weiner, & Morrison, 1978; Milton & Wiseman, 1999), but also positive results have been obtained using groups of respondents in ESP experiments (Barker, Messer, & Drucker, 1975; Carpenter, 1988, 1991; Dalkvist & Westerlund, 1998)—just as in the case of individual testing. More importantly, the number of group experiments on ESP reported thus far is too small to permit any well-founded evaluation of this type of experiment. In particular, almost no group experiments on telepathy have been reported.

One exception is a series of group telepathy studies that have been performed at the Department of Psychology, Stockholm University, since 1993, initiated by myself. Based on the idea that strong emotional messages (e.g., signals of danger) may be, for evolutionary reasons, easier to transmit telepathically than are more neutral messages (Moss & Gengerally, 1969), the studies have all been concerned with transmission of emotions, as evoked by slide pictures.

Each study comprised a series of repeated, identical experiments, with two groups of participants sitting in two adjacent rooms. In the first part of the experiment, participants in one of the two groups, the “sender” group, were presented with 15 emotionally positive and 15 emotionally negative pictures, one at a time. For each presented picture, participants in the other group, the “receiver” group, were to guess and note individually whether the picture was positive or negative. In a second part of the experiment, the two groups changed rooms and tasks: The former senders now served as receivers, and vice versa.

Hit rate, defined as the number or proportion of correct responses, was invariably used as the dependent variable in the data analyses. Hit rate was analysed as (a) a function of person or situational factors (e.g., belief in telepathy and the order in which the participant served as sender and receiver) and (b) a function of stimulus factors (e.g., rated characteristics of the target pictures). In recent studies, physical variables have also been examined, as detailed below.

As a first study in the present project (Dalkvist & Westerlund, 1998)—for an overview of the previous studies, see Table 1—five individual explorative studies were performed. They generated a large number of significant results (for example, an effect of belief in telepathy, with believers and nonbelievers performing worse than uncertain participants). Based on the significant results of the five explorative studies, which
apparently were unlikely to have been obtained by chance (even though we failed to support this judgment with any statistical measure), a set of hypotheses were subsequently tested in a comprehensive replication study (Westerlund & Dalkvist, 2004). The outcome of this study was not encouraging: None of the eight predictions was supported. The two remaining studies (Dalkvist & Westerlund, 2006; Dalkvist, Montgomery, Montgomery, & Westerlund, 2009) were more restricted in scope, as can be seen from Table 1.

The present study is based on data from several separate studies, four of which have remained unpublished until now. One is a new replication study. One motivation for conducting this study was to extend the amount of data to allow a highly reliable stimulus level analysis to be performed. The remaining three studies provided new scale descriptions of the target pictures.

Table 1
Summary of Previous Studies

<table>
<thead>
<tr>
<th>Main studies</th>
<th>Content</th>
</tr>
</thead>
</table>
| Dalkvist & Westerlund, 1998 | 1. Five different explorative substudies, resulting in several positive findings.  
                           | 2. Construction of psychological target picture scales.                                                                                  |
| Westerlund & Dalkvist, 2004 | Test of eight predictions from the initial study, none of which was supported.                                                          |
| Dalkvist & Westerlund, 2006 | Establishing the occurrence of a relationship between a sender/receiver order effect on performance, found in Westerlund & Dalkvist, 2004, and disturbances in the geomagnetic field. |
| Dalkvist, Montgomery, Montgomery & Westerlund, 2009 | Establishing reduced variability in hit rate in agreement with a sender/receiver order effect, found in Westerlund & Dalkvist, 2004. |

<p>| Scale construction studies                                                                 |</p>
<table>
<thead>
<tr>
<th>Study number</th>
<th>Rated picture features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All pictures: pleasant-unpleasant, involving, familiar, perceptible.</td>
</tr>
<tr>
<td>2</td>
<td>Positive pictures: calm, exciting.</td>
</tr>
<tr>
<td>3</td>
<td>Negative pictures: compassion-arousing, repulsive.</td>
</tr>
</tbody>
</table>

As indicated above, already at the beginning of the present project, the target pictures were rated on a set of psychological scales (see Table 1). Initially (Dalkvist & Westerlund, 1998), these scales showed some relationship to performance, but this relationship could not be replicated (Westerlund & Dalkvist, 2004).

These scales were judged to be inadequate, however. For one thing, their reliability could be questioned due to the small number of underlying ratings. Second, comparison of positive and negative pictures using the same scales for both types of pictures was not possible. To obtain more reliable and versatile psychological scales for use in future research, a new picture rating study was performed. This was done using participants in some of the experiments in the above-mentioned replication study (Westerlund & Dalkvist, 2004), without using the new scales in that study.

When at the start of our present project hit rate was chosen as the dependent variable, we assumed, quite naturally, that the receivers would follow the instructions and discriminate between positive and negative pictures—if they were able to show any discrimination at all. On closer reflection, however, this assumption is not at all obvious. After all, emotional reactions are complex phenomena, varying in at
least two global dimensions: pleasure-displeasure and arousal (perceived and/or physiological), the latter dimension exhibiting a more or less pronounced U-shaped relation to the former (Lang, Greenwald, Bradely, & Ham, 1993). Thus, the notion that some form of arousal rather than pleasure-displeasure (or some more specific pleasurable or displeasurable experience) might be transmitted from the senders to the receivers seemed to be conceivable and worth testing.

To allow for such comparisons, two studies were performed to extend the scale description of the targets to encompass not only psychological scales but also physiological ones. In one of the two physiological studies, electrodermal activity (EDA) was measured in response to the stimulus pictures, in the other, heart-rate (HR).

Increased EDA and decreased (!) HR enter as pivotal components in a pattern of various physiological reactions known as the orienting response, which is evoked by motivationally relevant stimuli that capture the individual’s attention (Sokolov, 1960). The two reactions are not perfectly correlated, however. While both of them tend to be somewhat stronger in response to negative as compared to positive stimuli, the difference between negative and positive stimuli is somewhat more pronounced for the decrease in HR than for the increase in EDA (Lang et al., 1993). Thus, the two measures complement each other to some extent.

The aim of the present study was to investigate whether participants could discriminate among the present target pictures and, if they could, to establish the conditions underlying this ability in terms of stimulus and moderator variables, using more powerful data than those previously used for similar purposes (Dalkvist & Westerlund, 1998; Westerlund & Dalkvist, 2004). Two sets of telepathy data were analysed, one referred to as the “old” and the other as the “new” data set. The old data set was obtained from the last two of the five initial explorative studies (Dalkvist & Westerlund, 1998), the best-controlled ones, and from the previous replication study (Westerlund & Dalkvist, 2004). The new data set was obtained from the new replication study, reported below. Only the new picture scales were used, not the original psychological scales.

Method

The Replication Study

Participants. Participants were 652 undergraduate students at the University of Stockholm, 459 females and 193 males, most of them at the Department of Psychology, who had chosen to participate in the study as part of course requirements.

As can be seen from Table 2, in terms of participant age and gender distributions, the present study was very similar to the corresponding previous studies, with a typical mean age of 26 or 27 years and a predominance of females, comprising about 70% of participants.

<table>
<thead>
<tr>
<th>Data collections</th>
<th>Mean Age</th>
<th>Females n (%)</th>
<th>Gender distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
<td>57 98 (69,7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>26.42</td>
<td>57</td>
<td>589 (69.7)</td>
</tr>
<tr>
<td>New</td>
<td>26.72</td>
<td>40</td>
<td>459 (70.4)</td>
</tr>
<tr>
<td>Total</td>
<td>26.55</td>
<td>57</td>
<td>1,048 (70.0)</td>
</tr>
</tbody>
</table>

As can be seen in Table 3, the present study was also similar to the previous ones with respect to number of experiments/sessions (the two parts of an experiment) and the distribution of participants across experiments/sessions. It should be noted, however, that the average number of participants per session was somewhat smaller in the new study than in the old ones. This was due to the occurrence of a larger number of small sessions in the new study than in the old ones, resulting from the students’ freedom to choose among different occasions for participating.
Table 3
Overview of the Old and the New Data Collections

<table>
<thead>
<tr>
<th>Data collections</th>
<th>N of experiments</th>
<th>N of sessions</th>
<th>Participants per session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>64(^a)</td>
<td>124</td>
<td>7.22</td>
</tr>
<tr>
<td>New</td>
<td>56(^b)</td>
<td>110</td>
<td>6.70</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>234</td>
<td>7.03</td>
</tr>
</tbody>
</table>

\(^a\) Four sessions were excluded due to apparatus failure.
\(^b\) Two sessions were excluded due to apparatus failure.

**Stimulus material.** The stimuli used were 30 slide pictures, 15 with positive motifs (such as nature pictures and pictures of happy people) and 15 with negative motifs (such as pictures of traffic accidents and starving children), collected specifically for the present project from various sources (IAPS, the frequently used emotion picture collection, did not exist when the project started). (For a detailed description of target pictures, see Dalkvist & Westerlund, 1998.)

**Procedure.** The new data collection was carried out exactly as in the first replication study (Westerlund & Dalkvist, 2004), but here one additional type of data was collected as well. In the new data collection, one of the two experimenters in the sender room registered any disturbance that occurred—coughs, rustling, scrapes, or any sound outside the sender room. Although the experimental rooms were soundproof, this sound registration was done to check whether any possible positive results could be accounted for by auditory (or vibrational) leakage, unconsciously perceived by the receivers and mediated by a correlation between disturbance and the two response alternatives (positive/negative picture).

In outline, the experiment was run as follows (for more details, see Westerlund & Dalkvist, 2004). When the participants arrived at the laboratory, they were randomly divided into two groups as equal in size as possible (the number of participants could be uneven): one sender group and one receiver group. The senders and the receivers were sequestered in two soundproof rooms, separated by one room. The two experimental rooms were connected to each other by a signal device: A lamp in the receiver room could be turned on and off from the sender room.

The slides were presented in random order for each group of senders, who sat in one or two rows in front of a screen. The senders’ only task was to look at the pictures and to “hold on to” the feelings evoked by the respective pictures as long as they were shown. The receivers, who sat in a circle with their backs to each other, were instructed to guess whether a given picture was positive or negative and to mark the chosen alternative on a response sheet (the receivers were forced to choose one of the two options). One of the (two) experimenters in the receiver room (there were also two experimenters in the sender room) watched the signal lamp, which was turned off when a new picture was shown to the senders, and reported this to the receivers. Each picture was shown for 20 s, with an interstimulus interval of about half a second.

When all 30 pictures had been shown, the participants changed rooms, and, as mentioned before, those who had served as senders in the first part of the experiment now served as receivers, and vice versa.

**Moderator variables.** As in Dalkvist and Westerlund (1998) and Westerlund and Dalkvist (2004), eight person/situation-related variables were examined as potential moderator variables. Two of them concerned belief in telepathy, one as measured before the experiment on a 3-point scale, and the other as measured after the experiment on a 7-point scale. Furthermore, the two available demographic variables, age and gender, as well as two different measures of response style: number of negative guesses and repetition avoidance, defined as the number of times the subject shifted from one type of response (“positive picture” or “negative picture”) to the other, were considered. The two remaining person/situation-related potential moderator variables were: number of receivers and the order of the two tasks, that as sender and that as receiver.
Moreover, as in Westerlund and Dalkvist (2004), two possible physical moderator variables were considered: (a) local sidereal time (LST)—an astronomical time and space measure, which is indirectly related to the magnitude of cosmic radiation that reaches the earth, and (b) fluctuations in the global geomagnetic field, as measured by the \(a_p\) index. For a large number of different studies, performed on the northern hemisphere, Spottiswoode (1997) found both of these measures to be systematically related to the effect size of the studies. (In a more recent paper [Sturrock & Spottiswoode, 2007], Spottiswoode repudiates his previous conclusions in favour of a lunar phase interpretation, but this interpretation has thus far failed to make a breakthrough.) In the present project, \(a_p\) index—but not LST—has been shown to be related to psi performance in two previous studies: namely, in Dalkvist and Westerlund (2006) and in Dalkvist et al. (2009). Based on previous findings showing LST around 13:30 to be psi-conducive (Spottiswoode, 1997), the LST scale was divided into two periods: a “good” period (coded as “1”), ranging from LST = 10:00 through LST = 16:00, and a “bad” period (coded as “0”), comprising all other times. (Spottiswoode’s psi-conducive window was, in fact, only 2 LST hours wide, but this small window would not have generated a sufficiently large amount of data).

The Rating Study

**Participants.** Participants were 66 undergraduate students at Stockholm University, 42 females and 24 males, with a mean age of 27.30 years. They had, in fact, been recruited to participate in the above-mentioned replication study (Westerlund & Dalkvist, 2004), where the present rating experiment served as a final, additional, session in some of the experiments.

Most of the participants were psychology students who chose to participate in the study as part of course requirements. All participants were informed beforehand that some of the pictures to be shown were very repulsive, and sensitive persons were recommended not to take part in the experiment.

**Stimulus material.** All the target pictures were the same as in the present replication study.

**Procedure.** The 30 target pictures were projected onto a white screen. The ratings were made in groups, with 10 or fewer participants in each group. The projector, being located behind the participants, was run automatically, by means of a timer. The target pictures were shown in a randomized order.

Each picture was rated on six different scales. Four of them measured purely emotional aspects of the pictures, namely how (a) pleasant/unpleasant, (b) involving, (c) compassion-arousing and (d) repulsive they were. The two remaining scales measured (e) how well known and (f) how perceptible (easy to understand) the motifs of the pictures were.

The participants were given a booklet of forms, each one containing the six scales, with one form for each target picture. For half of the subjects in a given session, the scales were written in the above order; for the other half, the order was reversed. Similarly, for half of the sessions, the target pictures were presented in one particular order, and for the other half in the reversed order. Each target picture was presented for 30 s. The judgments were made by drawing a vertical line on a 100-mm-long graphic scale with verbally anchored endpoints, representing extreme states of the experience to be judged (for example, “not at all disgusting” and “very highly disgusting”).

The EDA Study

**Participants.** Participants were 77 undergraduate students at the University of Stockholm, 43 females and 31 males, most of them at the Department of Psychology, who had chosen to participate in the study as part of course requirements. Due to disturbances in the equipment or loose electrodes, only 60 participants, 32 females and 28 males, entered into the final sample. These participants varied in age from 19 to 41 years, with a mean age of 24.72 years. All participants were informed beforehand that some of the pictures to be shown were very repulsive, and sensitive persons were recommended not to take part in the experiment. None of the participants had participated in any other study in the project.

**Stimulus material.** The target pictures were the same as in the present replication study.

**Apparatus.** An EDA monitor, with a software program for collecting and analysing data, manufactured by Biopac Systems, Inc., was used. The monitor (model mp 100A) and the program (AcqKnowledge III for the 100WS version 3.2) were linked to and installed in a Macintosh 6320 computer.
The electrodes used were of the EL204S-Ag/AgCl type. Isotonic paste served as electrolyte (0.5% NaCl/100 ml H$_2$O). The electrodes were fastened to the middle phalanges of the middle finger and the ring finger of the nondominant hand.

**Procedure.** The target pictures—the same as in the above study—were projected onto a white wall, in the same manner as in that study.

At most three participants took part in each measurement session. Each picture was presented for 20 s, with an interstimulus interval of about half a second. The total number of sessions was 35. Each session lasted for about 20 min, including instructions.

Three participant chairs, equipped with elbow rests, were located in a row in front of the projector.

The sequence of events during a session was as follows:

1. The participant(s) were asked to wash their hands with soap and water.
2. The participant(s) were given a complete description of the experiment, except for any expectation of the results or the connection to previous telepathy experiments. The participants were then informed that they were allowed to close their eyes if a picture was experienced as too unpleasant, and even to terminate the experiment.
3. The experimenter instructed the participant(s) on how to attach the electrodes by doing it on her own fingers and then letting the participant(s) do it themselves (without electrode paste).
4. The experimenter attached the electrode paste to the electrodes, and the participant(s) fastened the electrodes.
5. The participant(s) were instructed (a) to keep the hand with the electrodes as still as possible, but without concentrating too much on it, and (b) to put their arms as comfortably as possible on the arms of the chair. The participant(s) were also told not to speak to each other or to the experimenter during the experiment.
6. The light in the experimental room was turned off, and the projector was started. The first picture (a photo of a landscape) was not entered among the 30 “real” stimuli but was used only to accustom the participant(s) to the experimental situation, without the participant(s) knowing it.
7. After the last trial, the participant(s) were disconnected from the monitoring equipment and asked to fill out a short questionnaire, mainly concerned with demographic data.
8. The participant(s) were debriefed with respect to the purposes of the study. (In addition to testing the telepathy hypothesis, the data were also used in a study on gender differences in emotional reactions.)

**The HR Study**

**Participants.** Participants were 53 undergraduate students at the University of Stockholm, 37 females and 16 males, most of them at the Department of Psychology, who chose to participate in the study as part of course requirements. Due to missing data, only 50 participants, 36 females and 14 males, entered into the final sample. These participants varied in age from 19 to 39 years, with a mean age of 26.80 years. As before, all participants were informed beforehand that some of the pictures to be shown were very disgusting, and sensitive persons were recommended not to take part in the experiment. None of the participants had participated in any other study in the project.

**Stimulus material.** The target pictures were the same as in the present replication study.

**Procedure.** HR was measured by means of an electronic HR meter, manufactured by Polar Electro Oy. The HR meter (S610i™) consists of two parts: (a) a belt with heart-beat sensors, to be fastened around the chest, with the sensors adhering to the skin just below the breasts or the breast muscles, and (b) a “watch” for receiving and storing signals from the heart-beat sensors. Associated with the HR meter was a software program for data analysis.

There were at most four participants in each measurement session. Each picture was presented for 20 s, with an interstimulus interval of about half a second. Each session lasted for about 20 min, including instructions. There were 19 measurement sessions in all.
In order to avoid interference among the HR monitors, participants were spread out in the experimental room a couple of meters apart. In order to prevent females from becoming embarrassed when attaching the sensor belts, males were located in front of the females. On the experimenter’s command, the participant(s) started their “watches” just before the start of the experiment, and stopped them when the experiment was finished. Also, again on the experimenter’s command (saying “now”), the participant(s) pressed a button to register the point in time when a picture was exposed, and another button 7 s later to delineate the time interval for the measurement (the minimum time interval required for HR calculations was 5 s, that is, the same period as that used in measuring EDA, but with that short period one would have run the risk of obtaining periods that were too short to be analysed, due to delayed button presses when a picture was exposed.) The participant(s) could hear a “beep” when the button had been properly pressed.

The sequence of events was the same as for the EDA measurement study, except for the following steps:

1. The participant(s) were seated on chairs as described above.
2. Following the experimenter’s instructions, the participant(s) fastened the sensor belts as described above.
3. The participant(s) were told not to speak to each other or to the experimenter during the measurements and to concentrate on the pictures and not on handling the watch.
4. The participants were instructed how and when to start and stop their watches and how and when to make the time registrations, as described above.
5. A short trial run, without any pictures, was carried out.
6. The light in the experimental room was turned off, the participants started their “watches,” the experimenter started the projector, and so on.
7. After the last trial, the subjects took off their sensor belts and “watches” and filled out a short questionnaire, mainly concerned with demographic data.

Finally, the data were transferred to a computer for analysis.

**Results**

Data from all the individual studies involved in the present study—the four new studies reported above, the first replication study and the two best-controlled initial studies—were brought together with a view to answer this basic question: Could participants discriminate among the target pictures and, if they could, which were the critical mediating variables? Thus, the four studies are treated collectively and not one-by-one.

**Overall Analyses**

In the previous quest for relationships between picture characteristics and performance (Dalkvist & Westerlund, 1998; Westerlund & Dalkvist, 2004), a simple hit-rate measure was used (proportion of correct guesses). However, using this measure prevented us from comparing positive and negative pictures, because the measured performance would be affected by the participants’ preponderance to guess that any picture was, say, positive. To get around this problem, measures of relative hit rate with respect to the tendency to guess that any picture was positive or negative were used in the present study. Specifically, after having coded a participant’s hit in response to a given picture as “1” and a miss as “0,” each hit for a positive picture was divided by the participant’s total number of positive guesses and each hit for a negative picture by the participant’s total number of negative guesses, to transform any absolute hit into a relative hit. The individual relative hit rates for positive and negative pictures, respectively, were then calculated as the sum of all relative hits for positive pictures and the sum of all relative hits for negative pictures. The individual total hit rates, that is, the relative hit rates for both positive and negative pictures, were calculated as the mean of the relative hit rates for positive and negative pictures. To be used at the session level, each of the three sets of individual relative hit rates was averaged within sessions.
In running group experiments, one is most often confronted with a troublesome statistical problem, called the “stacking” problem: Due to the possible occurrence of dependency among participants’ responses in group testing (caused, for example, by the occurrence of a common response bias, such as a tendency to give one type of response at the beginning of a run and another type at the end of it), the statistical assumption of independent measures runs the risk of being violated, leading to deflated (or inflated, in cases of negative correlations) $p$ values (Thouless & Briar, 1970). In the present study, the risk of any stacking effect was eliminated by analysing data at the session level, in the form of group means. The reason why this procedure effectively eliminated any stacking effect is that participants’ responses to a given stimulus picture in a given session were independent of participants’ responses to the same stimulus picture in any other session because the stimulus orders were uniquely randomized.

In the statistical analyses, the new data were treated as an attempted replication of the old data, even though the old and the new data in reality were analysed simultaneously. Specifically, for each statistical test that was performed, the two data sets were submitted to a mini-Stouffer-$Z$ analysis ($N = 2$). In analogy with a regular meta-analysis, this was done not only with a view to testing the significance of the total results per se but also to test the replicability of the two separate data sets, by combining the Stouffer $Z$ analysis with assessment of the homogeneity of the data—in the present case, assessment of the difference between the old and the new data. It should be borne in mind, however, that such a replicability test is less stringent than performing a one-tailed significance test, and can only address the minimal criterion for claiming replicability.

### Picture Scales

Raw data were aggregated as follows in the three scaling studies.

In the rating study, for each scale and target picture, the arithmetical mean was calculated over all judges. In the EDA study, for each target picture, the mean of the response amplitude was calculated over a 5-s interval, starting from the exposure of the picture. These means were then transformed into $z$ values for each participant separately, to eliminate effects of individual response level differences. Finally, for each of the 30 target pictures, the mean $z$ value was calculated across the 60 (nonexcluded) participants.

In the HR study, by means of the program belonging to the measurement equipment, for each target picture, the mean response amplitude was calculated over a 7-s interval, starting from the exposure of the target picture. These means were then transformed into $z$ values for each participant separately, to eliminate effects of individual differences in response level. Finally, for each of the 30 stimulus pictures, the mean $z$ value was calculated across the 50 (nonexcluded) subjects.

For each of the three scaling studies, the interrespondent reliability was assessed using Cronbach’s alpha. In the rating study, reliability was assessed for each scale individually, yielding a mean alpha of .99 ($SD = .02$). For the EDA scale, Cronbach’s alpha was .51, and for the HR scale .24.

As expected from previous work (Lang et al., 1993), the EDA and HR scales were found to be negatively correlated. However, the correlation was small and far from significant, $r(28) = -.17, p = .36$, two-tailed.

Table 4 shows the Pearson correlations among the six rating scales. As can be seen from this table, most of the correlations are strong or very strong. As can also be seen from this table, the four emotional scales (1–4) are positively correlated with one another, but negatively correlated with the two, positively correlated, nonemotional scales.

<table>
<thead>
<tr>
<th>Scales</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasure-Displeasure (1)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassion (2)</td>
<td>.96**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repulsion (3)</td>
<td>.98**</td>
<td>.97**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involvement (4)</td>
<td>.71**</td>
<td>.85**</td>
<td>.77**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Familiarity (5)</td>
<td>-.74**</td>
<td>-.62**</td>
<td>-.75**</td>
<td>-.32</td>
<td>1</td>
</tr>
<tr>
<td>Perceptibility (6)</td>
<td>-.53**</td>
<td>-.40*</td>
<td>-.53**</td>
<td>-.03</td>
<td>.81**</td>
</tr>
</tbody>
</table>

* $p < .05$, two-tailed  
** $p < .01$, two-tailed
Table 5 shows the Pearson correlations between the two physiological scales and the six rating scales. As can be seen from this table, all of the four emotional rating scales are positively correlated with the EDA scale and negatively correlated with the HR scale, while the two nonemotional rating scales exhibit the reverse correlation pattern. It may also be noted that the correlations obtained for the HR scale are stronger than those obtained for the EDA scale, despite the fact that Cronbach’s alpha is higher for the latter scale than it is for the former.

Table 5

<table>
<thead>
<tr>
<th>Subjective scales</th>
<th>EDA</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pleasure-displeasure</td>
<td>.38*</td>
<td>-.47**</td>
</tr>
<tr>
<td>Compass</td>
<td>.36</td>
<td>-.43*</td>
</tr>
<tr>
<td>Repulsion</td>
<td>.39*</td>
<td>-.49**</td>
</tr>
<tr>
<td>Involvement</td>
<td>.30</td>
<td>-.42*</td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.31</td>
<td>.54**</td>
</tr>
<tr>
<td>Perceptibility</td>
<td>-.15</td>
<td>.32</td>
</tr>
</tbody>
</table>

* p < .05, two-tailed  
** p < .01, two-tailed

Given the strong correlations among most of the eight picture scales (the six subjective and the two physiological scales), all of them were combined into a composite scale. This was done by summing all of the original scales after having changed the signs of three of them—the HR, the familiarity, and the perceptibility scale—to make all scales positively correlated with one another. The resulting composite scale was interpreted as a bipolar scale measuring degree of unpleasant arousal, and was accordingly named a “negative arousal” (NA) scale. This summarizing scale was used to give an overall description of the target pictures.

Moderator Variables

As indicated by independent t tests, the old data studies had significantly larger values than the new study for preexperiment belief in telepathy, \( t(232) = 3.47, p = .001 \), two-tailed, as well as for postexperiment belief in telepathy, \( t(232) = 5.49, p < .001 \), two-tailed. They had also significantly larger values for number of receivers, \( t(232) = 3.33, p = .001 \), two-tailed, and on the \( a \) index, \( t(190) = 3.02, p = .003 \).

All Stimulus Pictures

The Pearson correlation between the old and the new data set in mean relative hit rate across all 30 stimulus pictures was \( r(28) = .31, p < .05, \) one-tailed, and the Spearman rank-correlation slightly higher: \( r(28) = .35, p = .03, \) one-tailed. Thus, as measured in terms of ordinary correlations, the two studies exhibited significant interstudy reliability with respect to relative hit rate.

The mean observed relative hit rate values for the old and the new data set, respectively, are given in Table 6, together with the results of one-sample t tests of deviation of mean observed values from mean chance expectation (MCE = .50) and a corresponding Stouffer Z analysis.

Table 6

<table>
<thead>
<tr>
<th>Data set</th>
<th>Mean observed relative hit rate</th>
<th>t</th>
<th>df</th>
<th>p (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>0.50</td>
<td>-0.02</td>
<td>123</td>
<td>.98</td>
</tr>
<tr>
<td>New</td>
<td>0.49</td>
<td>-1.49</td>
<td>109</td>
<td>.06</td>
</tr>
<tr>
<td>Both</td>
<td>Stouffer Z = 0.39, p = .35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As can be seen from this table, whereas the old data set did not show any deviation from MCE at all, the new one showed an almost significant tendency for relative hit rate to lie below MCE. Accordingly, the Stouffer Z test did not reach significance, as can be seen. The apparent difference in relative hit rate between the new and the old data was not significant, however. (The large difference between the two \( p \) values compared to the small difference between the two observed relative hit rate values is basically due to extremely small standard deviations, \( SD = .03 \) for the old data and \( .05 \) for the new.)

To test whether there was any significant overall effect of the stimulus pictures, one-way repeated measures ANOVA was applied to the old and the new data set, with the 30 stimulus pictures as the independent variable and mean relative hit rate as the dependent variable. As can be seen from Table 7, while the old data set did not show any effect at all, the new one approached—but did not reach—significance. Stouffer Z did not show any significant overall effect \( (Z = 1.28, p = .10) \).

Table 7
Results From One-way Repeated Measures ANOVA With All Stimulus Pictures as the Independent Variable and Relative Hit Rate as the Dependent Variable for the Old and the New Data Set and Results from a Stouffer Z Analysis

<table>
<thead>
<tr>
<th>Data set</th>
<th>( F )</th>
<th>df</th>
<th>( p ) (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>0.85</td>
<td>29/3567</td>
<td>.70</td>
</tr>
<tr>
<td>New</td>
<td>1.30</td>
<td>29/3161</td>
<td>.13</td>
</tr>
<tr>
<td>Both</td>
<td></td>
<td></td>
<td>Stouffer Z = 1.28, ( p = .10 )</td>
</tr>
</tbody>
</table>

Table 8 shows the Pearson correlations between mean relative hit rate and the eight original picture scales, as well as the corresponding composite (NA) scale. As can be seen from this table, a highly significant positive correlation was obtained for the EDA scale in the new data set, as well as a highly significant corresponding Z value. As can further be seen from Table 8, however, none of the remaining scales exhibited any significant effect. None of the differences between corresponding correlations for the two data sets was significant.

Table 8
Pearson Correlations Between Mean Relative Hit Rate and the Scales for the 30 Stimulus Pictures for the Old and the New Data Set and Results from Corresponding Stouffer Z Analyses

<table>
<thead>
<tr>
<th>Scales</th>
<th>Data set</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old ( r )</td>
<td>* ( p^* )</td>
<td>New ( r )</td>
<td>* ( p^* )</td>
<td>Both ( Z )</td>
<td>* ( p^* )</td>
<td></td>
</tr>
<tr>
<td>EDA</td>
<td>.30</td>
<td>.11</td>
<td>.52</td>
<td><strong>.003</strong></td>
<td>3.22</td>
<td><strong>.0006</strong></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.18</td>
<td>.35</td>
<td>-.18</td>
<td>.33</td>
<td>0.01</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>Pleasure-displeasure</td>
<td>.10</td>
<td>.60</td>
<td>.04</td>
<td>.82</td>
<td>0.53</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>Compass</td>
<td>.00</td>
<td>.10</td>
<td>-.03</td>
<td>.86</td>
<td>0.13</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>Repulsion</td>
<td>.05</td>
<td>.79</td>
<td>.05</td>
<td>.79</td>
<td>0.38</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>Involvement</td>
<td>-.17</td>
<td>.38</td>
<td>-.00</td>
<td>.99</td>
<td>0.62</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.17</td>
<td>.36</td>
<td>-.19</td>
<td>.31</td>
<td>1.35</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Perceptibility</td>
<td>-.26</td>
<td>.17</td>
<td>-.05</td>
<td>.79</td>
<td>1.16</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>.09</td>
<td>.65</td>
<td>.16</td>
<td>.39</td>
<td>0.93</td>
<td>.18</td>
<td></td>
</tr>
</tbody>
</table>

*two-tailed; bold represents \( p < .05 \)

Thus, considering all 30 stimulus pictures together, with the notable exception of the EDA scale, none of the eight original picture scales or the corresponding NA scale exhibited any linear relationship with mean relative hit rate.

Except for the EDA scale, however, rather than being linearly related to the picture scales, mean relative hit rate tended to exhibit U-shaped, or reversed U-shaped, relationships to these scales. For the old
and the new data set, the U- or reversed U-shaped trends could be summarized by a quadratic U-formed curve relating mean relative hit rate to the NA scale, corresponding to a Stouffer Z value of 2.93, \( p = .002 \). The curve was significant for the new data set, \( R = .51; F(2, 29) = 4.76; p = .02 \), two-tailed, but not for the old one, \( R = .23; F(2, 29) = 0.78, p = .48 \), two-tailed. As will be seen below, however, these U-formed relations are almost entirely attributable to a positive correlation between relative hit rate and the NA scale for the negative pictures, the positive pictures showing no clear correlation with the NA scale.

**Positive Versus Negative Pictures**

For the 15 positive pictures, the Pearson correlation between the old and the new data set in mean relative hit rate was \( r(13) = .05, p = .85 \), one-tailed, and the Spearman correlation was \( r_s(13) = .16, p = .58 \), one-tailed. For the 15 negative pictures, the Pearson correlation between the old and the new data set was \( r(13) = .52, p = .05 \), one-tailed, and the Spearman correlation was \( r_s(13) = .60, p = .02 \), one-tailed. Thus, whereas the positive pictures did not show any significant interstudy reliability at all, the negative pictures did so, the strength of the correlation between the two sets of measures exceeding .50, thus explaining more than 25% of the variance.

As can be seen from Table 9, for the positive pictures, a one-way repeated measures ANOVA did not show any effect at all. As can further be seen from this table, however, for the negative pictures, a nearly significant effect was obtained for the new data set, while no effect at all was obtained for the old one. As tested by a two-way repeated measures ANOVA, this apparent difference between the two studies was not statistically significant. As can further be seen from Table 9, however, despite the lack of significance in any of the separate analyses, Stouffer Z did reach significance.

### Table 9

<table>
<thead>
<tr>
<th>Data set</th>
<th>Positive pictures</th>
<th>Negative pictures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F )</td>
<td>( df )</td>
</tr>
<tr>
<td>Old</td>
<td>0.98</td>
<td>14/1722</td>
</tr>
<tr>
<td>New</td>
<td>1.00</td>
<td>14/1526</td>
</tr>
<tr>
<td>Both</td>
<td>Stouffer Z = 0.76, ( p = .23 )</td>
<td>Stouffer Z = 1.74, ( p = .04 )</td>
</tr>
</tbody>
</table>

*two-tailed; bold represents \( p < .05 \)

In Figure 1, applying a linear model, positive and negative pictures have been separately plotted against the NA scale for the old, the new, and the total data set. As can be seen from this figure, whereas the negative pictures exhibited a positive linear trend for each of the three data sets—although a stronger one for the new data set than for the old one—the positive pictures did not show any noteworthy linear (or other) trend (as will be seen below, the weak negative trends for the positive pictures are far from significant).

The Pearson correlations between the picture scales, including the NA scale, and mean relative hit rate for positive and negative pictures analysed separately are given in Table 10. As can be seen from this table, except for a moderately significant positive correlation for the HR scale in the old data set, no significant correlation was found for the positive pictures; accordingly, except for HR, no significant Z value was obtained for positive pictures. By contrast, as can also be seen from Table 10, several significant correlations, some of which are highly significant, were obtained for the negative pictures, but only for the new data set. No significant difference could be found between corresponding correlations in the old and the new study, however. The significant correlations in the new data set, as well as the corresponding correlations in the old one, are all positive. In agreement with the positive correlation obtained for the
NA scale, the original scales exhibiting positive correlations—the EDA, pleasure-displeasure and repulsion scale—can be said to characterize the pictures in terms of unpleasant arousal. The correlations obtained for the remaining original scales are all negative—both in the old and the new study, and with the exception of the HR scale, the corresponding scales describe the pictures in terms of “compassion,” “familiarity” or “perceptibility” rather than “arousal.” As can further be seen from Table 10, the scales exhibiting significant correlations were all associated with significant *Z* values.

In summary, the separate analyses of positive and negative pictures did not reveal any clear effect for the positive pictures, but did suggest that at least some of the negative pictures had been discriminated, at least in part based on the arousing aversive properties of the pictures. The significant *Z* values, along with the lack of any significant difference between corresponding correlations in the two studies and the overall agreement between their correlation patterns, show the two sets of correlations to be interrelated, even though no significant correlation was obtained for the old data set.

In view of these findings, it seemed appropriate to take an even closer look at the negative pictures.

### Negative Pictures

**Constructing a negative arousal discrimination scale.** With the aim of exploring how and to what extent participants tended to discriminate pictures associated with high negative arousal from pictures associated with low negative arousal, a negative arousal discrimination (NAD) scale was constructed. To this end, all negative pictures were divided into two categories based on a median split with respect to their NA scores. The sum of the relative hit rate values for the seven pictures with the highest NA scores and the sum of the relative hit rate values for the seven pictures with the lowest NA scores were then computed, for each participant separately. Next, the difference between the high arousal picture sum and the low arousal picture sum was computed. Finally, for the purpose of comparing session groups instead of single participants, the individual NAD values were averaged within session groups.

<table>
<thead>
<tr>
<th>Picture scales</th>
<th>Old data set</th>
<th>New data set</th>
<th>Both data sets</th>
<th>Type of picture</th>
<th>Old data set</th>
<th>New data set</th>
<th>Both data sets</th>
<th><em>r</em></th>
<th><em>p</em></th>
<th><em>r</em></th>
<th><em>p</em></th>
<th><em>Z</em></th>
<th><em>p</em></th>
<th><em>r</em></th>
<th><em>p</em></th>
<th><em>Z</em></th>
<th><em>p</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>EDA</td>
<td>.21</td>
<td>.46</td>
<td>.12</td>
<td>.66</td>
<td>.82</td>
<td>.21</td>
<td></td>
<td>.38</td>
<td>.16</td>
<td>.77</td>
<td>.001</td>
<td></td>
<td>3.31</td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>.56</td>
<td>.03</td>
<td>.07</td>
<td>.81</td>
<td>1.70</td>
<td>.04</td>
<td></td>
<td>-.08</td>
<td>.77</td>
<td>-.39</td>
<td>.151</td>
<td></td>
<td>1.21</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleasure-Displeasure</td>
<td>.36</td>
<td>.19</td>
<td>-.21</td>
<td>.46</td>
<td>.04</td>
<td>.34</td>
<td></td>
<td>.13</td>
<td>.64</td>
<td>.61</td>
<td>.016</td>
<td></td>
<td>2.03</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compassion</td>
<td>-.00</td>
<td>.10</td>
<td>-.26</td>
<td>.35</td>
<td>.67</td>
<td>.25</td>
<td></td>
<td>-.27</td>
<td>.33</td>
<td>-.16</td>
<td>.572</td>
<td></td>
<td>1.09</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repulsion</td>
<td>.36</td>
<td>.19</td>
<td>-.33</td>
<td>.24</td>
<td>.09</td>
<td>.46</td>
<td></td>
<td>.06</td>
<td>.83</td>
<td>.69</td>
<td>.005</td>
<td></td>
<td>2.13</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involvement</td>
<td>-.45</td>
<td>.09</td>
<td>.16</td>
<td>.57</td>
<td>.23</td>
<td>.41</td>
<td></td>
<td>-.19</td>
<td>.50</td>
<td>-.12</td>
<td>.678</td>
<td></td>
<td>0.77</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-.33</td>
<td>.23</td>
<td>.18</td>
<td>.51</td>
<td>.39</td>
<td>.35</td>
<td></td>
<td>-.19</td>
<td>.50</td>
<td>-.42</td>
<td>.118</td>
<td></td>
<td>1.58</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptibility</td>
<td>-.47</td>
<td>.07</td>
<td>.17</td>
<td>.54</td>
<td>.83</td>
<td>.20</td>
<td></td>
<td>-.22</td>
<td>.43</td>
<td>-.14</td>
<td>.620</td>
<td></td>
<td>0.90</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>-.09</td>
<td>.76</td>
<td>-.08</td>
<td>.78</td>
<td>.40</td>
<td>.34</td>
<td></td>
<td>.29</td>
<td>.29</td>
<td>.73</td>
<td>.002</td>
<td></td>
<td>2.93</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*two-tailed; bold represents *p* < .05*
Mean analyses. For each of the two studies, the mean of the NAD scores was compared to MCE (= 0) using a one-sample t test. As shown in Table 11, there was a positive deviation from MCE in both data sets, the deviation being significant for the new study, but not for the old one. The difference between the two studies did not reach but approached significance, \( t(232) = -1.72, \ p = .09, \text{ two-tailed} \). Thus, the significant Z value, given in Table 11, does not reflect the results in both studies, only the results in the new one.

Table 11

<table>
<thead>
<tr>
<th>Data set</th>
<th>MCE</th>
<th>Mean NAD score</th>
<th>t</th>
<th>df</th>
<th>p (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>0</td>
<td>0.002</td>
<td>0.35</td>
<td>123</td>
<td>.73</td>
</tr>
<tr>
<td>New</td>
<td>0</td>
<td>0.015</td>
<td>2.42</td>
<td>109</td>
<td>.02</td>
</tr>
<tr>
<td>Both</td>
<td></td>
<td>Stouffer Z = 1.83, \ p = .03</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bold represents \( p < .05 \)
In order to analyse the NAD scale more closely, the scale was split into its two components: the relative hit rate sum for the seven pictures with high NA values and the relative hit rate sum for the seven pictures with low NA values. This was done by averaging the respective individual relative hit rate sums within sessions.

Was one of the two sets of pictures—those with high and those with low NA values—more responsible for the positive deviations of the mean NAD scores from MCE shown in Table 11 than the other? To answer that question, the relative hit rate sums for the high and the low negative arousal pictures, respectively, were compared to MCE. As can be easily shown, MCE = 7 (number of low or high arousal pictures) / 30 (total number of pictures) = .23.

As can be seen from Table 12, there was a small and nonsignificant tendency for the high NA level pictures to have a mean relative hit rate sum above MCE, and a much stronger tendency for the low NA level pictures to have a mean relative hit rate sum below MCE, this deviation being significant in the new data set, but not in the old one. Thus, the significant positive NAD values shown in Table 11 resulted mostly from low NA level pictures having low relative hit rate sums rather than from high NA level pictures having high relative hit rate sums. In essence, this means that participants erroneously tended to guess that non-arousing negative pictures were positive instead of negative. In other words, whereas the high NA level pictures showed a (weak) tendency to evoke positive hit rates, the low NA level pictures were associated with so-called psi-missing, responses opposite in direction to the “correct” ones. This effect was clear-cut only in the new study, however. Thus, rather than demonstrating a common effect in the two studies, the significant Z value obtained for low NA pictures, shown in Table 12, only reflects the effect in the new study.

**Table 12**

*One-sample t Tests of Deviation of Observed Relative Hit Rate From MCE for Stimulus Pictures With High (Above the Median) and Low (Below the Median) NA Scores, Respectively, for the Old and the New Data Set and Results From Corresponding Stouffer Z Analyses*

<table>
<thead>
<tr>
<th>Data set</th>
<th>NA level</th>
<th>MCE for relative hit rate sum</th>
<th>Mean observed relative hit rate sum</th>
<th>t</th>
<th>df</th>
<th>p  (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>High</td>
<td>0.23</td>
<td>0.23</td>
<td>0.29</td>
<td>123</td>
<td>.78</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.23</td>
<td>0.23</td>
<td>-0.30</td>
<td>123</td>
<td>.77</td>
</tr>
<tr>
<td>New</td>
<td>High</td>
<td>0.23</td>
<td>0.24</td>
<td>0.95</td>
<td>109</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>0.23</td>
<td>0.22</td>
<td>-3.01</td>
<td>109</td>
<td>.003</td>
</tr>
<tr>
<td>Both</td>
<td>High</td>
<td>Stouffer Z = 0.83, p = .20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Stouffer Z = 2.19, p = .02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bold represents \( p < .05 \)

**NAD distributions.** Figure 2 shows the frequency distribution of the NAD values for each of the three data sets. Under the null-hypothesis, that is, the assumption that no relationship existed between negative arousal and relative hit rate, the NAD scores are expected to be symmetrically distributed, with MCE being equal to zero (the converse does not hold: that a distribution is symmetrical with a mean of zero does not exclude the occurrence of interaction effects, which then have to balance each other for the distribution to be symmetrical). As can be seen from Figure 2, however, at least as far as the shape of the distribution is concerned, this prediction is ostensibly violated: None of the three distributions seems to be symmetrical—not even approximately so; each distribution seems instead to be positively skewed, with a clearly visible prolonged “tail” pointing in the positive direction of the scale, the bulk of the distribution being located below zero.
This picture was confirmed and nuanced by numerical statistical data, which are shown beside the respective graphs in Figure 2. There are many methods for calculating a skewness measure. The most important ones are based on the second and the third moment about the mean, the latter moment indicating the basic degree of skewness. The present skewness measure is a sample version of such a measure ($G_z$), which is normally—or at least approximately normally—distributed (see Joanes & Gill, 1998). If $N$ exceeds 150, as is the case in the total data set, the present measure of skewness can—like any measure of skewness based on the second and the third moment—for all practical purposes be considered normal. For the present significance testing, an exact measure of the standard error of skewness, suggested by Cramer (1997), was adopted.

As can be seen from Figure 2, for all three data sets, the skewness value fell between 0.50 and 1.00—the lower and the upper limit, respectively, for a distribution usually being characterized as “moderately” skewed (a distribution with a skewness value above 1 is usually characterized as “strongly” skewed and a distribution with a skewness value below 0.50 as “slightly skewed” or “approximately symmetrical”). For the old and the new data set, the $z$ value and the corresponding $p$ value indicate that the skewness of each of the two distributions is clearly—though not remarkably—significant. For the total data set, however, the $z$ value and the corresponding $p$ value show the skewness of the distribution to be extraordinarily significant, with a $z$ value approaching 5 and a corresponding $p$ value of one in a million. Thus, the distribution for the total data set can apparently be characterized as skewed beyond any reasonable doubt. This conclusion was confirmed by a Stouffer $Z$ analysis, yielding a $Z$ value of 4.53 and a corresponding $p$ value of .000003, only marginally larger than that obtained in the skewness test.

**Moderator variables.** The fact that the distributions were found to be skewed, and not symmetric, indicates that session groups tended to respond differently to each of the two types of negative pictures—
those with high and those with low NA values. Hypothetically, this could be explained as resulting from interaction between negative arousal and some moderator variable(s). To discover whether one or more of the 10 potential moderator variables considered in the present study were involved, all of these variables were correlated with the NAD scale for the old and the new data set, respectively. As can be seen from Table 13, considering both the consistency across the two data sets, both of them yielding significant correlations, and the strength of the correlations, there was one—and only one—strong candidate: number of receivers. Thus, as the number of participants in the receiver group increased, the NAD value decreased significantly in both data sets, yielding a highly significant $Z$ value.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Old ($df = 122$)</th>
<th>New ($df = 108$)</th>
<th>Both ($df = 232$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$p^*$</td>
<td>$r$</td>
</tr>
<tr>
<td>Receiver order</td>
<td>-.02</td>
<td>.87</td>
<td>.20</td>
</tr>
<tr>
<td>Gender</td>
<td>.05</td>
<td>.62</td>
<td>.00</td>
</tr>
<tr>
<td>Belief in telepathy before the experiment</td>
<td>.11</td>
<td>.21</td>
<td>-.06</td>
</tr>
<tr>
<td>Belief in telepathy after the experiment</td>
<td>.05</td>
<td>.58</td>
<td>-.02</td>
</tr>
<tr>
<td>$N$ of receivers</td>
<td>-.23</td>
<td>.01</td>
<td>-.21</td>
</tr>
<tr>
<td>Age</td>
<td>.01</td>
<td>.88</td>
<td>.12</td>
</tr>
<tr>
<td>Repetition aversion</td>
<td>-.13</td>
<td>.14</td>
<td>.08</td>
</tr>
<tr>
<td>$N$ of negative guesses</td>
<td>-.18</td>
<td>.05</td>
<td>-.15</td>
</tr>
<tr>
<td>LST (Good – Bad)</td>
<td>-.12</td>
<td>.17</td>
<td>.12</td>
</tr>
<tr>
<td>$ap$-index (Ln)</td>
<td>-.00</td>
<td>.97</td>
<td>-.26</td>
</tr>
</tbody>
</table>

*two-tailed; bold represents $p < .05$

In the following analyses involving moderator variables, attention will be focused on number of receivers. Even though they did not reach significance in both data sets, the two variables number of negative guesses and $ap$ index will also be considered, as these two variables exhibited significant $Z$ values, thus fulfilling a minimal requirement for claiming replicability.

Of these two variables, the more interesting one is perhaps $ap$ index, as activity in the geomagnetic field has been linked to psi performance in several previous studies (e.g., Arango & Persinger, 1988; Berger & Persinger, 1991; Haraldsson & Gissurarson, 1987), most often with a low level of geomagnetic activity apparently being psi-conducive. In the afore-mentioned study by Spottiswoode (1997), a negative correlation was found between $ap$ index and psi performance—but only in an apparently psi-conducive LST window of 2 hours around 13:30 LST. The possibility that such an interaction between $ap$ index and LST occurred also in the present study was tested in relation to the NAD scale. The result was negative, however.

As can be seen from Table 14, the positive correlation between number of receivers and the NAD scale was mainly due to a positive correlation between the relative hit rate sum and number of receivers for the low NA level pictures, and only to a lesser extent to a negative correlation between the relative hit rate sum and number of receivers for the high NA level pictures. As can also be seen from Table 14, there is good general agreement between the correlation patterns in the old and the new study, as reflected by a highly significant $Z$ value for the low NA level pictures and a less significant $Z$ value for the high NA level pictures.
Performance in Group Telepathy Experiments

Table 14
Pearson Correlation Between N of Receivers and Relative Hit Rate Sum for High and Low NA Level Pictures for the Old and the New Data Set and Results From Corresponding Stouffer Z Analyses

<table>
<thead>
<tr>
<th></th>
<th>Old (df = 122)</th>
<th></th>
<th>New (df = 108)</th>
<th></th>
<th>Both (df = 232)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p*</td>
<td>r</td>
<td>p*</td>
<td>Z</td>
<td>p</td>
</tr>
<tr>
<td>High</td>
<td>-.10</td>
<td>.27</td>
<td>-.15</td>
<td>.12</td>
<td>1.86</td>
<td>.032</td>
</tr>
<tr>
<td>Low</td>
<td>.27</td>
<td>.003</td>
<td>.19</td>
<td>.05</td>
<td>3.54</td>
<td>.0001</td>
</tr>
</tbody>
</table>

*two-tailed; bold represents p < .05

The exact relationships among the three current variables: NA level, relative hit rate sum and number of receivers are shown in Figure 3. As can be seen, for each of the three data sets, the low NA level pictures show a steep—essentially linear—increase in relative hit rate sum in small session groups as the number of receivers increases. In contrast, starting from a higher relative hit rate level, the high NA level pictures show a (less steep) decrease in relative hit rate sum with increasing number of receivers in small session groups. At a group size of about six participants, the two curves converge, and with increasing group size both curves level off.

Figure 3. Relative hit rate sum for high and low NA level pictures, respectively, plotted against number of receivers for the old, the new, and the total data set. The horizontal lines indicate the performance value expected by chance (MCE = .233).
Together the two curves clearly demonstrate the nature of the current effect: a failure on the part of the receivers in small, as opposed to large, groups to identify negative pictures having low negative arousal scores as negative, that is, psi-missing.

As mentioned earlier, the skewness of the distributions of NAD scores shown in Figure 2 could be expected to result from interaction between relative hit rate and one or more moderator variables. If this is true, given the negative correlation between number of receivers and the NAD scale, eliminating the effect of number of receivers should lead to a reduction—or even elimination—of the skewness of the NAD distributions. A reduction of the effect of number of receivers on the NAD scale was, in fact, accomplished by extracting the residuals of the NAD scores with respect to number of receivers using linear regression. In line with the expectation, the distributions of the residuals obtained were found to be considerably less skewed than the original NAD distributions, as can be seen from Table 15. As can also be seen from Table 15, however, the distributions were still positively skewed, and significantly so for the new and the total data sets. As can further be seen, however, by also entering the two other moderator variables that are significantly correlated with the NAD scale in the total data set—number of negative guesses and $ap$ index—into the regression analysis, the skewness of the distributions was reduced additionally, resulting in nonsignificant skewness values for both the old and the new data set. Thus, the present analysis provided strong support for the interaction interpretation of the skewed NAD scale distributions, even though the skewness could not be fully eliminated by removing the effects of the three most influential moderator variables.

### Table 15

**Skewness Tests for NAD Score Residuals Obtained From Linear Regression Analyses With NAD Score as Dependent Variable and Moderator Variables as Independent Variables**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Statistics</th>
<th>Original data</th>
<th>Residuals for</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$N$ of receivers</td>
</tr>
<tr>
<td>Old</td>
<td>Skewness</td>
<td>0.64</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>$z$</td>
<td>2.95</td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td>$p(124)^*$</td>
<td><strong>0.002</strong></td>
<td>0.06</td>
</tr>
<tr>
<td>New</td>
<td>Skewness</td>
<td>0.34</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>$z$</td>
<td>3.19</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>$p(110)^*$</td>
<td><strong>0.0007</strong></td>
<td>0.003</td>
</tr>
<tr>
<td>Total</td>
<td>Skewness</td>
<td>0.76</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>$z$</td>
<td>4.75</td>
<td>2.85</td>
</tr>
<tr>
<td></td>
<td>$p(234)^*$</td>
<td><strong>0.000001</strong></td>
<td><strong>0.003</strong></td>
</tr>
</tbody>
</table>

*one-tailed; bold represents $p < .05$

### Control Analyses

Four different control tests were performed to check whether the major positive results obtained could be explained away as artifacts. The major positive results were considered to be: (a) the positive correlation between mean relative hit rate and the NA scale for negative pictures, (b) the positive skewness of the NAD scale distribution for negative pictures, and (c) the correlation between
the NAD scale and number of receivers for negative pictures. Four possible “natural” explanations of these findings were tested:

1. The positive results are methodological artifacts resulting from averaging individual data into session data. Although it is hard to imagine how such artifacts could have arisen, I recalculated the major positive results using individual data instead of group data, and compared these new results with the original ones. The individual level correlations between picture scales and mean relative hit rate for negative pictures showed the same pattern as the group level correlations, shown in Table 10. The strength of the correlations between mean relative hit rate and the NA scale was somewhat reduced, however, resulting in a reduction of Stouffer $Z$ from 2.93, $p = .002$, to 2.14, $p = .02$.

A more pronounced weakening of the results was obtained for the skewness measure of the NAD scale distributions, shown in Figure 2. Most notably, for the whole data set, the skewness value was reduced from the extremely significant ($p = .000001$) value of 0.76 to the much smaller and much less significant ($p = .004$) value of 0.18.

A reduction in the strength of the results was also obtained for the correlations between number of receivers and the NAD scale, shown in Table 13, resulting in a corresponding diminishing of Stouffer $Z$ from 3.35, $p = .0004$, to 3.20, $p = .0007$.

How should this weakening of the positive results be explained? Before attempting to answer this question, it should be emphasized that the positive results were not eliminated by shifting from the session level to the individual level, but were merely weakened. This fact excludes the possibility that the session level results were mere artifacts caused by averaging individual data. Indeed, the occurrence of weaker results at the individual level than at the session level is what is to be expected given the negative relationship between number of receivers and the NAD scale. Thus, in accordance with this relationship, when session averages are calculated from individual data, participants in larger groups, showing only small or no arousal effects at all, will get a heavier weight than do participants in smaller groups, who do show such effects, simply because participants in larger groups are more numerous than those in smaller groups. Conversely, at the session level, the impact of the large number of participants belonging to large groups becomes relatively small, because all groups are treated equally regardless of their number of participants.

2. The positive results were artifacts of transforming absolute into relative hit rate values. To test this interpretation, the same positive results as considered above were recalculated using absolute instead of relative hit rates as input data. Only a negligible change followed. Instead of being reduced, the Stouffer $Z$ reflecting the correlations between relative hit rate and the NA scale was, in fact, slightly increased: from 2.93, $p = .002$, to 3.01, $p = .001$; the skewness value for the NAD scale distribution for the total data set was reduced from $G_1 = 0.76$, $p = .000001$, to $G_1 = 0.67$, $p = .00001$; and for the correlations between the NAD scores and number of receivers, Stouffer $Z$ was reduced from 3.35, $p = .0004$, to 3.20, $p = .0007$.

The next interpretation is specifically concerned with the significant skewness values of the NAD scale distributions (Figure 2), especially the remarkably significant value in the case of the total data set.

3. The skewness of the distribution of NAD values was due to asymmetrically distributed extreme values. Any measure of the skewness of a distribution is sensitive to asymmetrically distributed extreme values, especially outliers. To test whether the occurrence of asymmetrically located extreme values could account for the small $p$ values associated with the skewness values of the NAD scale distributions—especially the remarkably small $p$ value in the case of the total data set—the tails of the distributions were truncated by moving extreme values toward the centre of the distribution. This was done by categorizing the NAD values using open end-categories before the skewness analysis started. Specifically, for a given data set, the original NAD scale was transformed into a 10-point scale with equal intervals, except for the two end-categories, which had equal frequencies (5) in the case of the total data set. As can be seen from Figure 4, these open categories effectively eliminated any possible effect of extreme values. With the original NAD values being replaced by corresponding category values (1–10), new skewness values and their associated $z$ and $p$ values were calculated. Somewhat surprisingly, as can be seen from Figure 4, the effect of these recalculation was to strengthen (!), instead of weaken, the skewness indicators, yielding, for example, a $p$ value close to 1 in 10 million (!) for the total data set. This strengthening of the skewness indicators was most likely due to
reduction of sampling errors resulting from the transformation of the original NAD data into categorical data, thus counteracting the skewness-reducing effect of truncating the distribution. It may also be noted that, whereas the distribution for the new study previously was somewhat more skewed than that for the old one, the two current distributions are practically identical.

4. Positive findings were due to disturbances in the sender room. As may be recalled, during the new data collection, occasional disturbances in the sender room (scraps, coughs, and other sounds), which, at least in theory, might have been unconsciously perceived by the receivers, were registered by one of the two experimenters in the sender room. Typically, a few (minor) disturbances occurred during a session ($Md = 3$), only 23 of the 110 sessions (20.9%) being completely free from any disturbances. There was no significant correlation between number of negative guesses and number of disturbances as calculated across sessions, $r(107) = -.015$, $p = .88$, two-tailed. Likewise, there was no significant correlation between number of disturbances in a session and the NAD scores, neither for small session groups, defined as $n < 6$, $r(45) = .16$, $p = .28$, two-tailed, nor for all session groups regardless of number of receivers, $r(108) = -.057$, $p = .55$, two-tailed. Hence, at least as far as the new data are concerned, the major findings of the present study could not be due to auditory (or vibrational) perceptual leakage.

Discussion

Major Results

Implicitly, there was one general assumption behind the present study: Considering the whole set of stimulus pictures, performance was expected to be related to at least one of the eight picture scales, or to all of them in combination. This general assumption was supported by two findings: First, as measured by correlations between relative hit rate among the 30 stimulus pictures for the old and the new study,
there was significant interstudy reliability. Second, the NA scale, summarizing the eight original scales, was significantly related to relative hit rate (by a quadratic function) as was the EDA scale (by a linear function), as indicated by Stouffer Z.

However, the strongest results were obtained through a systematic explorative procedure, aimed at unveiling the roots of the above findings. A hierarchical strategy was followed in this search, comprising two major steps. In the first step, positive and negative pictures were analysed separately. Whereas no clear effect was found for the positive pictures, several different analyses suggested that participants did discriminate among negative pictures. In the second step, the negative pictures were analysed in more detail, as discussed below.

Not surprisingly, in the many studies on the dimensional structure of emotions that have been performed during the twentieth century, a dimension of pleasure versus displeasure has repeatedly appeared as the most salient one. However, in the present study, the distinction between pleasant and unpleasant emotions did not appear to be related to psi performance. Instead a dimension of negative arousal, varying above all among the negative pictures, seemed to be critical.

As far as the negative pictures are concerned, the systematic quest for any relationship between pictures and receivers’ responses resulted in three major findings:

1. Participants discriminated between two types of negative pictures. As related to relative hit rate, two types of negative stimulus pictures could be discerned: arousing and nonarousing ones. In terms of the eight original picture scales, the arousing pictures were characterized by particularly high levels of EDA and repulsiveness, and the nonarousing pictures by high familiarity or compassion levels. Both types of pictures tended to give rise to relative hit rate values that deviated from MCE—but in different directions. Thus, the predominantly arousing pictures tended to be associated with responses above MCE, that is, hits, and predominantly compassion-eliciting or familiar pictures with responses below MCE, that is, psi-missing. This latter deviation was the most pronounced one, and was significant for the new data set, although not for the old one.

While there was no evidence that participants had distinguished between positive and negative emotions, or between different positive emotions, the results suggested that participants did discriminate between two types of negative emotions: repulsion and compassion. Indeed, these two emotions are quite distinct—phenomenologically and physiologically as well as behaviourally. Thus, while compassion feels “soft,” is physiologically characterized by parasympathetic rather than sympathetic autonomic reactions, and tends to evoke approaching, care-giving behaviour, repulsion feels “hard” and “tense,” is dominated by sympathetic rather than parasympathetic activity, and evokes avoidance or withdrawal behaviour (see, e.g., Rollenhagen, 1990). There is thus a tangible basis for any discrimination between repulsion and compassion, making the present results potentially understandable in psychological as well as physiological terms.

2. The NAD scale distribution was found to be positively skewed. Under the general null hypothesis that no telepathic communication or any other psi phenomenon occurred in the present study, a distribution of NAD scores is expected to be completely symmetric (under the null hypothesis, there is no logical reason why the sum of the relative hit rate values for the seven high-arousal pictures should be larger than the sum of the relative hit rate values for the seven low-arousal pictures). At odds with this prediction, however, for each of the three data sets, the distribution of NAD scores was found to be significantly positively skewed. According to standard criteria, the degree of skewness obtained for the three data sets could neither be characterized as very high nor as very small, but as moderate. However, in the case of the total data set, apparently due to the large $N$ (= 234), a significance test showed the skewness to be extremely significant, with a $z$ value of 4.75, corresponding to a $p$ value of one in a million. With respect to the amount of skewness, the NAD scale distributions for the old and the new data set agreed very well with that obtained for the total data set. However, although the skewness was clearly significant for both the old and the new data set, the corresponding $z$ values (2.95 and 3.19, respectively) were considerably smaller than that obtained for the total data set, and, although small, the corresponding $p$ values (.002 and .0007, respectively) were not exceptionally small.

Different control tests failed to explain away the skewness of the NAD scale distributions as due to any bias resulting from averaging individual data, the occurrence of asymmetrically distributed extreme-values, or the use of relative, instead of absolute, hit rate as the measure of performance.
But what about the extremely small $p$ value obtained in the case of the total data set? Is this value really trustworthy?

First of all, there is no reason to dispute the validity of the present skewness test. For example, there was no underlying assumption that was violated, such as the requirement of a sufficiently large number of cases (at least in the case of the total data set, rather than being small, the sample was large according to established criteria), or independent measures (the different session groups could not in any way have affected each other’s responses, or have been affected in the same way in responding to any given target picture, as the pictures were presented in uniquely randomized orders).

At least in theory, however, there is one possible remaining “natural” explanation for the extremely small $p$ value in the case of the total data set: Such small $p$ values may be unstable and therefore unreliable. In the present case, this would mean that, if the study had been only slightly different, the $p$ value of one in a million might instead have become, say, one in a thousand or (why not) one in a hundred million. Whether or not this general statement about small $p$ values being associated with a high degree of uncertainty is true (which, in fact, may be disputed), the statement seems at least not to be true in the present study. The $p$ value for the total data set is thus perfectly consistent with the combination of the $p$ value obtained for the old data set ($p = .002$) and that obtained for the new one ($p = .0007$), neither of which can be characterized as extremely small, the product of the two $p$ values being equal to $p = .0000014$—almost exactly the same $p$ value as that obtained for the total data set.

3. The skewness of the NAD scale distribution could largely be accounted for in terms of interaction between negative arousal and number of receivers. The skewness of the NAD scale distributions strongly suggests that some interaction had occurred between negative arousal and some moderator variable(s) (though not necessarily any of those considered in the present study). This interpretation was supported by a correlation analysis: For each of the three data sets, a significant negative correlation was found between the NAD scale and one of the 10 potential moderator variables of the present study: number of receivers. This relationship was for the most part attributable to a positive correlation between relative hit rate and number of receivers for the low negative arousal pictures, rather than a negative correlation between relative hit rate and number of receivers for the high negative arousal pictures.

The negative correlation between the NAD scale and number of receivers could account for a substantial part of the skewness of the NAD scale distributions, as shown by eliminating the effect of number of receivers using linear regression—but not all of it. A further reduction of the skewness was obtained by entering two additional moderator variables into the regression analysis: number of negative guesses and $ap$ index. But the resulting distributions were still positively skewed. Perhaps some further, unknown, moderator variable, or some combination of the present ones, is needed to fully account for the skewness of the present NAD distributions. Additional potential moderator variables that should be tested in future research are the person-describing variables (age, gender, belief in telepathy and the two response style measures—number of negative guesses and repetition aversion) as applied to the senders instead of the receivers. Preliminary tests seem, in fact, to indicate that particular characteristics of the senders (the two response style variables) may be more important than the corresponding characteristics of the receivers.

So far in this discussion, the relation between the NAD scale and the size of the receiver group has been described as linear. However, the relationship was, in fact, only approximately linear. More specifically, whereas the NAD value decreased progressively with increasing group size for small receiver groups (< six receivers), there were no apparent differences in relative hit rate for large receiver groups (> five receivers). In other words, there was an upper limit around five receivers for any difference in relative hit rate between the two types of pictures to appear. This finding can apparently explain why the old data set in general exhibited weaker results than the new one, as the session groups tended to be smaller in the new data set than in the old one.

The difference between small and large receiver groups now discussed can be said to support a compromise between the common view, mentioned in the introduction, that group testing is inefficient in producing positive psi results, and the opposite view that group testing is equally, or more, efficient than individual testing in this respect. Thus, according to the present study, individual testing is more psi-conducive than group testing, but group testing will also do, provided that the groups are not too large. It remains to find out, however, why the size of the receiver group would be critical.
A comment should also be devoted to the significant negative correlation between the NAD scale and \(ap\) index for both data sets together. This relationship confirms previous suggestions that fluctuations in the geomagnetic field may be related to psi performance, even though Spottiswood’s specific finding of a psi-conducive window around 13:30 LST could not be replicated. Specifically, according to the present study—as well as several previous experimental studies (see, e.g., Berger & Persinger, 1991), a low level of geomagnetic fluctuations is psi-conducive, suggesting that telepathy somehow is transmitted by, or otherwise dependent on, electromagnetic fields. Perhaps the old, but nowadays less popular idea among parapsychologists that telepathy—if it does exist—is an electromagnetic phenomenon should be reconsidered.

**Critical Methodological Discussion**

The really critical question is, of course, if there is any reasonable “natural” explanation for the present positive results. During the initial studies in the present project, every effort was made to eliminate any perceptual leakage or any other experimental flaw, such as involuntary communication between the experimenters in the sender room and those in the receiver room—including a telepathy-based experimenter effect (the two experimenters in the sender room didn’t know which picture the senders were presented with at any given trial). Also, by testing the randomness of the random orders of the stimulus pictures, we made sure that they were free from any bias. And before the first replication experiment (Westerlund & Dalkvist, 2004) started, we were entirely convinced that any experimental error, however far-fetched it seemed, could be excluded—except for one remaining possible error. In spite of the fact that the senders were instructed to be silent, and the fact that the two experimental rooms were sufficiently well sound isolated to prevent any normal sound in the sender room to be heard in the receiver room (loud screams—but not ordinary talk—could be heard from one room to the other), there were occasional minor disturbances in the sender room, which at least in theory, could have been unconsciously perceived by the receivers. Hence, to the extent that there was any correlation between these disturbances and characteristics of the stimulus pictures, the disturbances could have affected the receivers’ responses systematically. However, the control test made in this study, based on noted disturbances in the sender room, excludes this possibility. It is thus very unlikely that the present positive results can be explained by any experimental error whatsoever.

But what about statistical errors? The most critical question is perhaps whether the extremely small \(p\) value of one in a million that was obtained in testing the skewness of the NAD scale distribution for the total data set can be taken seriously, or resulted from some statistical bias, such as one or several unfulfilled statistical assumptions or the instability of extremely small \(p\) values. As argued above, however, as far as I can see, there is no reason to dispute the validity of the extremely small \(p\) value now being considered.

This extremely small \(p\) value is not unique to the skewness test, however. A very small \(p\) value can also be obtained by combining larger significant, independent, \(p\) values. For example, this holds for a combination of the \(p\) value obtained for the correlation between the NAD scale and number of receivers: \(.0003\), and the \(p\) value for Stouffer \(Z\) for the two correlations between the NA scale and relative hit rate: \(.002\). This yields a combined \(p\) value of \(.0003\times.002 = .0000006\).

One major reason why the extremely small \(p\) values just discussed are important in interpreting the present positive results is that they argue against the likelihood that these results were due to positive selections from post hoc tests: It does not seem reasonable to argue that a \(p\) value of one in a million or thereabout can be found merely through such selections. In principle, the selection hypothesis can be tested by adjusting the \(p\) values for number of tests being performed using a Bonferroni-like method (multiplying the original \(p\) value by the number of tests). Unfortunately, however, in most practical situations, this cannot be done in a rigorous manner, for two major reasons. One is that the decision of how many tests to include in the analysis is often completely arbitrary. The other is that the method is only valid if the component analyses are independent, an assumption that is grossly violated in most actual research cases. Nevertheless, we can still say that, according to the Bonferroni logic, more than 10,000 independent tests are required before we can conclude that the most significant results may be due to multiple testing. Thus it is clear that these results cannot be solely accounted for by positive selection.
In the present paper, instead of attempting to adjust \( p \) values, an alternative approach has been taken: to consider the new study as a replication of the old one and to compare the results of the two studies, both with respect to similarities and with respect to differences. Arguably, to the extent that such similarities or differences can be established, the selection interpretation diminishes in credibility, even though no proper prediction testing is being performed.

As pointed out above, the existence of a general agreement between the two studies was suggested by the interstudy reliability analysis for the whole set of pictures. This conclusion was strengthened and qualified by the results of corresponding analyses for the negative and positive pictures separately, showing significant interstudy reliability for the negative pictures, but not for the positive ones. Thus, for both the whole set of stimulus pictures and for the negative pictures, the Pearson correlation as well as the Spearman rank-order correlation between the two data sets with respect to relative hit rate was significant.

More specifically, the major positive findings discussed above were all characterized by at least some agreement between the old and the new study, as indicated by Stouffer \( Z \) analyses in conjunction with assessment of differences between the old and the new study. Particularly good agreement was obtained (a) for the skewness of the NAD distribution and (b) for the correlation between number of receivers and the NAD scale: in both cases, not only was the Stouffer \( Z \) value highly significant, but the results were also independently significant in the old and the new study. In the case of the relation between relative hit rate and the NA scale, the picture is somewhat less clear. Thus, neither this scale nor any of its component scales exhibited a significant correlation with relative hit rate in the old study, only in the new one. Nevertheless, the highly significant Stouffer \( Z \) value, the agreement between the correlation patterns in the old and the new study, as well as the lack of any significant difference between corresponding correlations in the two studies point to a common effect. The above agreements indicate that at least some of the findings were replicable: To some extent, results from the new study could thus be “predicted” from the old one, and, conversely, results from the old study could be “retrodicted” from new results.

Throughout the analyses, there was a tendency for the new study to yield clearer results than the old one, even though most of the noted differences were not significant, but only suggestive (the only exception is the correlation between the \( ap \) index and the NAD scale, which was significantly negative in the new study but zero in the old one). This tendency for the new study to yield clearer results than the old one may be explained by some minor methodological differences between the two studies. One obvious possible explanation is the tendency for the session groups to be smaller in the new study than in the old one.

**Concluding Remarks**

It should not be forgotten, however, that previous attempts to predict future results in the present project have failed blatantly, which should be taken as a warning against drawing too hasty conclusions. However, the results are certainly sufficiently strong and interesting to warrant future tests. Unfortunately, though, in the current version, the present study is probably too time- and resource-consuming for any researcher to be willing or able (it is no easy task to recruit almost 1,600 participants) to perform a more exact replication of the study. Luckily, however, a reduced version of the present study is accessible and would do. Thus, taking advantage of the negative correlation between the NAD scale and number of receivers, a small number of receivers in each session group would suffice to demonstrate an effect. It would also be sufficient to use only negative pictures, provided they include both repulsive and familiar or compassion-evoking pictures. Mere conceptual replications based on the concept of negative arousal would also do. For example, some previous data (e.g., ganzfeld data associated with emotional target descriptions) might be re-analysed in terms of hit rate as related to some available measure of arousal. Also, by amplifying critical picture characteristics, for example, by using film clips instead of slide pictures, the effects might be strengthened, thereby reducing the need for large amounts of data.

In any case, the present study would seem to provide a potential recipe for bringing forth telepathic communication or some other psi phenomenon.
References


Acknowledgments

The research reported in this paper was supported by a grant from the John Björkheim Memorial Foundation. I am grateful to my friends and colleagues Gergő Hadlaczky and Joakim Westerlund for their theoretical and practical support during the progress of this work. I am also grateful to Eric Nissen for helpful comments on an earlier version of this paper.

Abstracts in Other Languages

French

LES PERFORMANCES DANS LES EXPERIMENTATIONS DE TELEPATHIE EN GROUPE EN FONCTION DES CARACTERISTIQUES DE L’IMAGE CIBLE

RESUME : Nous avons analysé des données de communication télépathique en groupe d’émotions, ainsi qu‘évoquées par des diapositives, en cherchant à identifier les caractéristiques critiques des images. Les performances furent associées avec 6 échelles psychologiques liées aux images ainsi que deux échelles mesurant la tendance des images à évoquer des réactions électrodermales (EDA) et cardiaques (HR), respectivement. Toutes les échelles ont été regroupées en une échelle composite, reflétant l’excitation négative (NA). En se basant sur cette échelle, une échelle de discrimination de l’excitation négative (NAD) fut construite. Un ancien ensemble de données, obtenues auprès de 845 participants, et un nouvel ensemble obtenu auprès de 652 participants, étaient analysés à la fois séparément et conjointement, avec comme mesure de performance le taux de succès relatif (taux de succès en fonction de la tendance à réagir). La fiabilité entre études fut établie de manière significative. Les participants discriminent entre 2 types d’images négatives : celles avec un haut potentiel d’excitation négative et celles avec un faible potentiel. Pour les images négatives, l’échelle NAD fut significativement déviée positivement, avec une valeur $p$ remarquablement petite pour l’ensemble des données ($p = 0.00000$). Cette déviation pouvait largement – mais pas totalement – être expliquée en termes d’une corrélation négative significative ($p = 0.0003$) entre les scores NAD et le nombre de récepteurs. Les découvertes majeures se trouvent clairement dans le nouvel ensemble de données. Diverses explications « naturelles » des résultats positifs sont discutées et éliminées car improbables. Une expérimentation simplifiée est suggérée pour de futures réplications.

Spanish

RENDIMIENTO EN EXPERIMENTOS GRUPALES DE TELEPATÍA EN FUNCIÓN DE LAS CARACTERÍSTICAS DEL OBJETIVO

RESUMEN: Los datos sobre comunicación grupal telepática de emociones, evocada por imágenes en diapositivas, fueron analizados con el fin de identificar las características importantes de las imágenes. Los resultados se analizaron con 6 escalas psicológicas de las imágenes, así como con 2 escalas de medición de la tendencia de las imágenes para evocar respuestas electrodermales y de tasa cardíaca, respectivamente. Agrupó las 8 escalas en una escala compuesta para reflejar excitación negativa (EN). En base a esta escala, construí una escala de discriminación de excitación negativa (EDE). Analice la tasa de éxito relativo (tasa de éxito en relación con la tendencia de respuesta) como medida de ejecución en una muestra vieja de datos con 845 participantes, y una nueva con 652 participantes, tanto separada como conjuntamente. Encontró confiabilidad significativa entre los estudios. Los participantes discriminaron entre 2 tipos de imágenes negativas: las de alto y de bajo potencial de excitación negativa. Para las imágenes negativas, la escala EDE estuvo sesgada positivamente, con un valor muy bajo de $p$ para todos los datos ($p = 0.000001$). Esta asimetría puede explicarse en gran medida en términos de una correlación negativa significativa.
(p = 0.0003) entre las puntuaciones de EDE y el número de receptores. Los hallazgos más importantes tendieron a ser más claros en los datos nuevos que en los antiguos. Discuto varias explicaciones “naturales” de los resultados y las descarto como poco probables. Describo un experimento simplificado para futuros intentos de replicación.

**German**

LEISTUNG BEI EXPERIMENTEN ZUR GRUPPENTELEPATHIE ALS FUNKTION DER MERKMALE DES ZIELBILDES

AN ANOMALY OF AN ANOMALY: INVESTIGATING THE CORTICAL ELECTROPHYSIOLOGY OF REMOTE STARING DETECTION

BY IAN S. BAKER AND PAUL STEVENS

ABSTRACT: If there is evidence of an overall effect of remote staring detection, then theoretically there should also be evidence of electrophysiological processing of this information in the brain. A series of three experiments examining the potential electrocortical correlates of remote staring detection are presented, followed by a fourth experiment to examine a potential artifact. The first experiment provided an initial exploration of this effect, finding primarily that “remote staring detection” has no evident time-locked processing associated with it on its own but rather acts upon other processes occurring at the same time. The second experiment provided evidence that this effect is not related specifically to face processing but can impact on other forms of processing as well. The third experiment uncovered evidence of a potential artifact that could explain the “remote staring effect,” which is verified in the final experiment. The overall results are discussed in light of an interesting and subtle psychophysics luminance effect that could potentially have an impact upon a wide variety of experiments that employ event-related measures of electrocortical processing.

Keywords: remote staring detection, electroencephalography, global field power, psychophysics, luminance artifact

Remote staring detection has been defined as “... the purported ability to detect when one is being watched or stared at by someone situated beyond the range of the conventional senses.” (Braud, Shafer, & Andrews, 1993a, p. 391). Remote staring detection involves the measurement of behavioural or physiological reactions in starees when stared at by a starer, even though it should be impossible for the starees to know through any conventional sensory means that the starer is staring at them at any particular moment. Belief in this phenomenon as an everyday experience is considerably widespread, with incidences of belief ranging from approximately 70% to 94% of the populations sampled (Braud et al., 1993a; Braud, Shafer, & Andrews, 1993b; Coover, 1913; Cottrell, Winer, & Smith, 1996; Rosenthal, Soper, & Tabony, 1994; Sheldrake, 2003; Thalbourne & Evans, 1992). Over the past 100 years there have been several attempts to examine these anecdotal experiences and beliefs under controlled conditions. The earliest research in this area used relatively simple and direct behavioural measures that demonstrated an evolution of methodological sophistication over time as greater controls over extraneous variables were introduced (Coover, 1913; Poortman, 1959; Titchener, 1898; Williams, 1983). The introduction of the use of electrodermal activity (EDA) as a measure of autonomic nervous system (ANS) activity and as a potential indicator of a “fight-or-flight” response to being stared at remotely was a significant methodological development. This was particularly the case when the EDA method was combined with the use of CCTV systems to separate the starer and staree (Braud et al., 1993a, 1993b). Collectively referred to as the “EDA-CCTV” studies (Baker, 2005), several researchers found interesting results utilizing this method, including potential skeptic-believer experimenter effects (Schlitz & LaBerge, 1994; Schlitz, Wiseman, Watt, & Radin, 2006; Watt, Schlitz, Wiseman, & Radin, 2005; Watt, Wiseman, & Schlitz, 2002; Wiseman & Schlitz, 1997, 1999; Wiseman & Smith, 1994, 1994). A meta-analysis (Schmidt, Schneider, Utts, & Walach, 2004) of the 15 EDA-CCTV experiments that had been conducted at that time found a small but significant effect ($d = .13, p = .01$), suggesting evidence that requires further investigation.

This was the primary objective of the research presented in this paper. Firstly, previous EDA-CCTV methods were expanded to include central nervous system (CNS) activity. It would be expected that, if this phenomenon is genuine, then any stimulus processing or awareness of a remote stare should result in corresponding activity in the brain. Secondly, it was important to embed the potential effect within a wider theoretical framework. Assuming that remote staring detection is producing brain activity as the information is processed, does this processing follow similar systems to those that have already been identified in cognitive neuroscience; for example, the processing of faces and/or the gaze of others?
The significance of various forms of eye-based nonverbal communication in humans has been long established in the social psychology literature (e.g., Argyle & Cook, 1976; Ellsworth, Carlsmith, & Henson, 1972; Kirkland & Lewis, 1976). The human eye has the largest ratio of exposed, white sclera to dark iris compared to any other primate (Kobayashi & Kohshima, 1997; Riccardelli, Baylis, & Driver, 2000), which appears to aid humans in being particularly sensitive to the detection of gaze and its direction (Itier, Van Roon, & Alain, 2011). The impact of the gaze of another also elevates electrodermal measures of arousal (Helminen, Kaasinen, & Hietanen, 2011; Leavitt & Donovan, 1979; McBride, King, & James, 1965; Nichols & Champness, 1971; Strom & Buck, 1979), which neatly correlates with the EDA-CCTV measures of remote staring detection mentioned previously. The impact of face and gaze processing on electrical brain activity has been well-studied over the past decade and a half. Different components have been identified, but there appears to be significant activation surrounding the $T_5$ (or $P_7$) and $T_6$ (or $P_8$) electrodes, particularly a negative component in the right (i.e., $T_6/P_8$) hemisphere at approximately 170 ms after stimulus onset (Allison, Puce, Spencer, & McCarthy, 1999; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Carmel & Bentin, 2002; McCarthy, Puce, Belger, & Allison, 1999), with the subtle differences between eyes-only and face processing as a whole being under debate (Farroni, Csibra, Simion, & Johnson, 2002; Itier, Latinus, & Taylor, 2006; Taylor, Itier, Allison, & Edmonds, 2001; Watanabe, Kensaku, & Ryusuke, 2002). It is becoming increasingly apparent that the processing of eye and gaze stimuli may represent a core substrate of social cognition (George & Conty, 2008; Itier & Batt, 2009).

A series of studies is reported here that examined the potential existence of electrocortical correlates of remote staring detection, and the potential association of such correlates with more conventional forms of eye or face perception. The method and analysis procedure for each of the four experiments was identical unless otherwise noted. A series of planned and post hoc analyses was conducted on the electroencephalographic (EEG) data from each experiment, including event-related potentials (ERPs), topographical analyses, fast Fourier transforms (FFTs), evoked and induced event-related band power (ERBP) and partial-least squares (PLS). Additionally, extensive analysis was conducted on the skin conductance (SC) data and questionnaire data. However, due to space limitations only the primary ERP analysis (as reflected by the global field power [GFP] analysis), the frequency analysis of the entire stimulus epoch for Study 2, and the analysis of skin conductance are reported here (see Baker, 2007 for more details of the other analyses).

This first study had two main objectives: (a) was there any evidence of global electrocortical processing of remote staring detection, and (b) did any such processing vary in any way when administered at the same time as more conventional face processing? Such an approach was designed to act as an initial step in a program of research to examine if remote staring detection and face/gaze processing were potentially utilizing the same brain processes.

**Experiment 1**

This experiment examined if remote staring detection was revealed in global electrocortical processing, and if this processing may have an interaction with more conventional forms of face processing. Accordingly, participants were exposed to four different conditions: (a) conventional face processing, (b) a remote stare, (c) both face and remote stare processing together, and (d) no stimulus as a control condition.

**Method**

**Participants.** Twenty participants (7 males and 13 females) took part in this experiment with an average age of 26.0 years (range: 21–41 years). The participants were not paid and were selected using an opportunity sampling method. The majority of the participants were right-handed (two were left-handed). This and all of the following experiments received ethical approval from the Ethics Committee of the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh, and conformed to BPS and APA ethical guidelines. All participants provided informed consent.
Materials, equipment, and procedure. The experiment broadly followed the procedure outlined in previous remote staring detection and direct mental interaction with living systems (DMILS) experiments (Braud & Schlitz, 1991; Braud et al., 1993a, 1993b), with additional elements due to the use of more complex electrophysiological methods. The schematic of the setup of the experimental equipment is shown in Figure 1. Participants were initially oriented in the testing laboratory and asked to complete three questionnaires: a general demographics questionnaire, a 23-item Self Consciousness Questionnaire (Burnkrant & Page, 1984; Fenigstein, Scheier, & Buss, 1975; Mittal & Balasubramanian, 1987), and a 20-item nonclinical paranoia questionnaire (Fenigstein & Vanable, 1992). The analysis of the questionnaire data is not reported in this paper due to space limitations (see Baker, 2007, for more details). The skin conductance (SC) electrodes were then applied to the medial phalanges of digits 2 and 3 of the participant’s nondominant hand (as per the guidelines set out by Fowles et al., 1981). Unfortunately, due to equipment failure the skin conductance data for Experiment 1 was corrupted and could not be analysed. The EEG cap was then fitted according to the 10-20 system (Jasper, 1958). The recording electrodes were as follows: Fp1, Fp2, F7, F8, F3, F4, Fz, T7, T8, C3, C4, Cz, T3, T4, T5, T6, P3, P4, Pz, O1, O2, A1, A2, VEOG+, VEOG−, HEOG+, HEOG−, and ground (GNR). The EEG cap was connected to the NeuroScan NuAmps Amplifier, which was connected to the EEG recording computer. Triggers for the EEG were sent via an optically isolated connection from the experimental computer. The starer and staree were physically isolated from one another in two rooms that were 25 meters apart, and the staree was in a sound-attenuated room.

![Figure 1. Schematic of the equipment setup for all three experiments.](image)

During the experiment, the participant was exposed to four separate conditions. Each condition was repeated 48 times in a pseudorandomised and counterbalanced order that was automatically changed by the computer for each participant. During certain conditions starees were presented with a static picture of the starer on the screen in front of them, at other times it was blank. In addition, during these times starees may also be stared at remotely by the starer via the computer-controlled CCTV system, depending on the condition. These four conditions are summarised in Table 1.
The participants’ EEG was recorded at 500 Hz sample rate, with a bandpass filter at 0.5 Hz (high pass) and 100 Hz (low pass) with a 50-Hz notch filter. Each condition lasted for 5,000 ms followed by a 5,000-ms rest period.

Results and Discussion

The EEG data were preprocessed to remove muscle and ocular artifacts, epoched and averaged into event-related potential (ERP) data for each condition (-100-ms to 500-ms epochs). The data from all of the electrodes were then summarised using GFP (Lehmann & Skrandies, 1980; Nunez & Srinivasan, 2006; Skrandies, 2002). GFP can be expressed as the following (from Lehmann & Skrandies, 1980):

$$GFP = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} (U_i - U_j)^2}$$

This formula represents the root-mean-square deviations between all electrodes (i.e., for each of the voltages $U$ for an $i \times j$ array of $n$ electrodes) for each time point (based on Skrandies, 1995). Essentially, the GFP values represent the spatial standard deviation between all electrodes over time. It is a highly robust measure that uses the data from all of the recording electrodes, and it is also independent of the reference site (Lehmann & Skrandies, 1980). It is partly due to these reasons that it is the primary method for peak identification as recommended in the Society for Psychophysiological Research recording standards and publication criteria (Picton et al., 2000). Peak detection was used as the main measure here as it is discreet and easily definable, which is vital for the a priori definition of the measurement of a phenomenon that has never been examined in this way before. As a function of the calculation, all values are positive. The GFP data for all participants and for all four conditions can be seen in Figure 2. Two primary temporal peaks were identified as being of interest: 134 ms and 222 ms, and an additional slower peak from 378–500 ms. Unlike the two earlier peaks, the 378–500-ms epoch did not have a clear and distinct peak but is of interest due to the novel nature of the stimuli. Due to the duration of this, an area-under-the-curve (as opposed to a mean) measure was used in this analysis. Shapiro-Wilk analyses revealed that the data were not normally distributed and therefore nonparametric analyses were conducted. Alpha levels were corrected using a modified Bonferroni procedure (Keppel, 1982; Russel, 1990), giving an $\alpha_{MB} = .01$ in this instance.

The analyses focused upon demonstrating a potential remote staring detection effect by comparing the Remote Stare and Control conditions, and the potential relationship between face processing and remote staring detection (by comparing the Face and the Face + Remote Stare conditions). The analysis demonstrated that there was a significant difference between the peak GFP amplitudes for the Face and the Face + Remote Stare conditions for both the 134-ms ($z = -2.88, p = .004$) and 222-ms peaks ($z = -2.43, p = .01$), but not for the 378–500-ms time period ($z = -0.04, p = .97$). There were no significant differences between the peak GFP amplitudes for the Remote Stare and Control conditions for the different time points (134 ms: $z = -0.60, p = .55$, 222 ms: $z = -0.30, p = .77$, 378–500 ms: $z = -0.22, p = .82$). In addition to the
findings above, there was also a clear and highly significant effect of face processing when the Face and Control conditions were compared (134 ms: \( z = -3.81, p < .001 \), 222 ms: \( z = 3.92, p < .001 \), 378–500 ms: \( z = -3.88, p < .001 \)), although these results need to be treated with caution due to the considerable difference in stimulus types.

![Global field power (GFP) results from all 20 participants for all conditions in Experiment 1.](image)

**Figure 2.** Global field power (GFP) results from all 20 participants for all conditions in Experiment 1.

The results suggested that there was a significant effect of remote staring detection on global measures of brain activity. However, this effect was only present when the remote staring stimulus was administered in conjunction with the face perception stimulus, where the remote staring stimulus apparently significantly reduced peak GFP associated with face processing. There was no significant difference between the global brain activity of the remote staring stimulus on its own compared to the control stimulus.

**Experiment 2**

The results from the first study were curious as they suggested that remote staring detection does not cause any apparent distinct brain activity in its own right but rather appears to be dependent upon other processing that is occurring concurrently. However, it was unclear from the findings of the first study if this finding represented a unique relationship between face/gaze processing and remote staring detection, or if remote staring detection acted upon any concurrent process.

In order to examine this, the second experiment exploited the debate concerning cortical domain specificity associated with face processing compared to the processing of other objects. Since faces are processed differently than most objects, some researchers have suggested it is due to the specific social importance of faces to humans (e.g., Bentin & Carmel, 2002; Carmel & Bentin, 2002; Itier et al., 2006; Kanwisher, 2000). Therefore, by examining how remote staring detection may impact the processing of both faces and objects, it was possible to discern if the process involved in remote staring detection was a face-specific interaction or a more general interaction involving a wider range of concurrent stimulus processing.
Method

Participants. Twenty participants (7 males and 13 females) took part in this experiment with an average age of 25.3 years (range: 20–38 years). The participants were paid for taking part and were all staff or students at the University of Edinburgh. All but one of the participants were right-handed.

Materials, equipment, and procedure. Apart from relatively minor equipment upgrades, all of the equipment was identical to that used in Experiment 1. The relevant EEG and skin conductance electrodes were attached in the same manner. The same personality questionnaires were administered.

The overall procedure was the same as that used in Experiment 1, except that the conditions that the participant was exposed to were different. Each condition was repeated 60 times in a pseudo-randomised and counterbalanced order. Apart from the rest periods, the participants were presented with either a static picture of the starer on the screen in front of them or by a picture of a chair from the International Affective Picture Set (IAPS) database. A chair was used in order to reflect the maximum degree of processing differences between faces and objects (Itier & Taylor, 2004). In addition, during these times the staree may also have been stared at remotely by the starer via the computer-controlled CCTV system, depending upon the condition. These four conditions are summarised in Table 2.

<table>
<thead>
<tr>
<th>Action of starer</th>
<th>Staree’s screen</th>
<th>Staree’s screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face displayed</td>
<td>Face + Remote Stare condition</td>
<td>Object + Remote Stare condition</td>
</tr>
<tr>
<td>No Remote Stare</td>
<td>Face condition</td>
<td>Object condition</td>
</tr>
</tbody>
</table>

The participants’ EEG was recorded at 500 Hz (32-bit) sample rate, with a high-pass filter at 0.5 Hz and no low-pass filter (system maximum range was 262.5 Hz) and no notch filter. Each condition lasted for 5,000 ms followed by a 5,000-ms rest period.

Results and Discussion

Again, the EEG data were preprocessed to remove muscle and ocular artifacts, epoched, and averaged into event-related potential (ERP) data for each condition. These epochs were slightly longer than in Experiment 1 in order to encapsulate any potentially later effects (-100 ms to 800 ms). Global field power was the main measure used, and two temporal peaks were identified as being of interest: 150 ms and 208 ms. The GFP data for all participants and for all four conditions can be seen in Figure 3. Shapiro-Wilk analyses revealed that the data did not violate any assumptions of normality and so parametric analyses were conducted.

Separate 2 × 2 (image type × remote staring manipulation) repeated measures ANOVAs were conducted on the two peaks of interest. The initial 150-ms component demonstrated a significant effect for remote staring processing, $F(1,19) = 6.95, p = .02$, but no significant difference between face and object processing, $F(1,19) = .18, p = .68$, and no significant interaction effect, $F(1,19) = .002, p = .97$. The second (208-ms) component mirrors these findings, with a significant effect for remote staring processing, $F(1,19) = 23.23, p < .001$, no significant difference between face and object processing, $F(1,19) = .45, p = .51$, and no significant interaction effect, $F(1,19) = .02, p = .90$. Additional analyses (See Baker, 2007, for more details) revealed that the differences in face and object processing were broadly localised to the right temporal lobe region (i.e., $P_8/T_6$) as expected (Eimer, 2000; Itier, & Taylor, 2004).

One potential issue with the ERP/GFP analyses is that they examine only a small part of the data; only the first 800 ms of a 5,000-ms epoch. As the phenomenon under investigation has not been examined in this way previously, it was possible that a “remote staring effect” may be noted over a longer
duration. However, ERP/GFP analyses are not suited for this. In order to examine (a) the relationship with alpha activity, and (b) a longer time duration, a post hoc analysis of global alpha activity (using fast-Fourier transforms) for all four conditions over the 5-s stimulus period (divided into the averaged activity for each second) was examined. A $4 \times 5$ (conditions $\times$ time [seconds]) repeated measures ANOVA with Greenhouse-Giesser correction revealed no significant differences between the alpha activity of the different conditions, $F(1.077, 20.456) = 0.82$, $p = .38$, and no significant effect of time, $F(1.158, 21.999) = 1.73$, $p = .20$, and no significant interactions, $F(1.069, 20.304) = 0.96$, $p = .34$. This indicated that for the most dominant frequency band in the evoked domain there was no remote staring detection effect over a longer duration; it was only evident in the peak elements of the GFP (see Baker, 2007 for additional post hoc analyses).

![Figure 3. Global field power (GFP) results from all 20 participants for all conditions in Experiment 2.](image)

Finally, the skin conductance data for each of the 60 administrations of each stimulus for each person were averaged and compared for each condition. Analysis indicated no significant differences in skin conductance between the Face condition and Face and Remote Stare condition ($z = -0.58$, $p = .56$) or between the Object condition and the Object and Remote Stare condition ($z = -1.85$, $p = .07$). However, skin conductance responses to stimuli can rapidly habituate in as little as 2 to 8 stimulus administrations (Dawson et al., 1990). In order to investigate this, two post hoc analyses examined the averaged skin conductance responses to the first 16 (similar to previous skin conductance studies into remote staring detection: e.g., Schlitz & LaBerge, 1997), then first 8 administrations of each stimulus, similar to above. However, none of these comparisons approached significance (see Baker, 2007, for more details).

The results suggest that remote staring detection has an effect upon the global processing of both faces and objects—increasing the GFP in both cases—and does not appear to be a face-specific effect. In conjunction with the results of the first study, it suggests that remote staring detection apparently does not have an electrocortical processing in its own right, but rather acts upon any concurrent processing.

The lack of any processing of remote staring detection on its own and the fact that the impact of remote staring detection on faces reversed between the two studies (in the first study it reduced the peak GFP, in the second study it increased the peak GFP) was concerning. This reversal might be due to the subtle
methodological differences between the two studies. In the first experiment, the randomisation sequence resulted in participants effectively being presented with an image at fairly random intervals, whereas in the second experiment the image presentation was very regular. This may have altered alpha activity generation between the experiments and produced different effects (Shaw, 2003). Alternatively, it may have revealed a potential artifact that caused this significant “remote staring effect.” The third experiment was designed to replicate the previous effects and test for the possibility of an artifact.

Experiment 3

The third experiment replicated the conventional face processing condition and the face and remote stare condition used in the two previous experiments in order to examine the reversal of the effects between experiments one and two in more detail. In addition to this, the third experiment also examined the possibility that the effect of the remote staring detection was an artifact. This was done by simply removing the remote staring stimulus altogether for half of the experiment, but otherwise conducting the experiment as before. The rationale behind this was simple: remove the remote stare, and—if it was a genuine effect—this should remove the effect itself.

Method

Participants. Twenty participants (10 males and 10 females) took part in this experiment with an average age of 27.8 years (range: 18–50 years). The participants were paid 5 pounds for taking part and were all staff or students at the University of Edinburgh. All but two of the participants were right-handed.

Materials, equipment, and procedure. All of the EEG and skin conductance equipment, the other experimental hardware, and the questionnaires were the same as for the last experiment. The overall procedure was the same as for the last two experiments, apart from some minor alterations due to the type of conditions that the participants were exposed to in this experiment. In order to examine the effect of the removal of the starer on the remote staring effect, a pseudorandomised and counterbalanced split-half design was used. For 50% of the sessions, the starer was physically present for the first half of the session and absent for the second half of the session. For the other 50% of the sessions, this was reversed. The order in which this occurred was randomised (without replacement) by an independent party (the second author), and the experimenter (the first author) was not aware of the order of any session prior to the session beginning. Within each half of the session, the order of the Face or the Face + Remote Stare conditions was also pseudorandomised and counterbalanced. This resulted in four conditions that are summarised in Table 3.

<table>
<thead>
<tr>
<th>Action of starer</th>
<th>Starer’s screen (and remote stare manipulation)</th>
<th>Face Only displayed</th>
<th>Face + Remote Stare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starer Present</td>
<td>Face (Starer Present) condition</td>
<td>Face + Remote Stare (Starer Present) condition</td>
<td></td>
</tr>
<tr>
<td>Starer Absent</td>
<td>Face (Starer Absent) condition</td>
<td>Face + Remote Stare (Starer Absent) condition</td>
<td></td>
</tr>
</tbody>
</table>

The participants’ EEG was recorded using the same parameters as the second experiment.

Results and Discussion

As in the two previous experiments, the EEG data were preprocessed to remove muscle and ocular artifacts, then epoched and averaged into event-related potential (ERP) data for each condition (epochs of
115
Anomaly of an Anomaly: The Cortical Electrophysiology of Remote Staring Detection

-100-ms to 800-ms duration). Once again, global field power was the main measure used, and two temporal peaks were identified as being of interest: 120 ms and 174 ms. The GFP data for all participants and for all four conditions can be seen in Figure 4. Shapiro-Wilk analyses revealed that the data did not violate any assumptions of normality and so parametric analyses were conducted.

Separate 2 × 2 (presence of starer × remote staring manipulation) repeated measures ANOVAs were conducted on the two peaks of interest. The initial 120-ms component demonstrated a significant effect for both remote staring processing, \( F(1,19) = 10.18, p = .005 \), and for the presence of a starer, \( F(1,19) = 12.01, p = .003 \), but no significant interaction effects, \( F(1,19) = 0.01, p = .87 \). The second (174 ms) also suggested a significant effect for remote staring processing, \( F(1,19) = 54.89, p < .001 \), but no significant effect for the presence of a starer, \( F(1,19) = .03, p = .87 \), and no significant interaction effects, \( F(1,19) = 1.72, p = .21 \).

![Figure 4](image-url)

**Figure 4.** Global field power (GFP) results from all 20 participants for all conditions in Experiment 3.

However, these results can only be understood to their fullest extent by examining them with paired-sample *t* tests. Two comparisons for each peak of interest were conducted. The first compared the Face (Starer Present) and the Face + Remote Stare (Starer Present) conditions. As the starer was physically present during each of these conditions, this test is equivalent to the Face and the Face + Remote Stare comparisons that were conducted in the first two experiments, and therefore ostensibly tests for the impact of remote staring detection on the global processing of faces. The second comparison examined the differences between the Face + Remote Stare (Starer Present) and the Face + Remote Stare (Starer Absent) conditions. The test between these two conditions more clearly examines the impact of physically removing the remote starer from the experiment than the analyses above. Significant results here would suggest the remote starer is important to this effect, nonsignificance would support the existence of a potential artifact.

The initial 120-ms peak demonstrated a significant difference between the Face (Starer Present) and the Face + Remote Stare (Starer Present) conditions, \( t(19) = -2.16, p = .04 \), but it did not suggest a significant difference between the Face + Remote Stare (Starer Present) and the Face + Remote Stare (Starer Absent) conditions, \( t(19) = -1.21, p = .24 \). The findings for the second peak (174 ms) mirror these findings, with a significant difference between the Face (Starer Present) and the Face + Remote Stare
Finally, a 2 × 2 ANOVA (remote staring × presence of starer) analysis of the averaged skin conductance responses for the 60 administrations of the different conditions suggested that there were no significant effects for remote staring detection, $F(1, 19) = 1.3, p = .26$, or for the effect of the presence or absence of a starer, $F(1, 19) = 0.03, p = .86$, or any significant interaction between these two factors, $F(1, 19) = 0.004, p = .95$.

These results suggested that the “remote staring effect” was potentially caused by some form of experimental artifact. As the experiment was computer-controlled and the conditions were the same in all conditions—with the exception of whether the camera feed to the starer’s monitor was masked or not—it suggested that there was some alteration of the images that the staree was looking at (and therefore the electrocortical processing associated with them). As the image that the staree was presented with was the same computer file for all conditions, the image presentation for the different conditions needed to be examined in case the physical properties of the image were somehow changing between conditions and the participants were reacting to this change.

**Experiment 4**

In order to examine the physical properties of the image, a sensitive photodiode was used in order to examine the luminance levels of the image presentation in the different conditions. As the Starer Present and Starer Absent conditions from Experiment 3 were equivalent from an equipment perspective, only the Face (Starer Absent) and Face + Remote Stare (Starer Absent) conditions were used for the comparison. This was important because in the Face conditions the camera feed was masked, and in the Face + Remote Stare it obviously was not.

**Method**

**Materials, equipment, and procedure.** The experimental setup was as similar as possible to the procedure of the third experiment. The only main difference was that there was a photodiode reacting to the images on the staree’s screen rather than a participant. The photodiode (BPW21: OSRAM Opto Semiconductors) was positioned 150 mm away from the center of the staree’s screen. The photodiode had a relative spectral sensitivity that is close to that of the human eye. It was connected to a Gould Advanced Digital Storage Oscilloscope OS4000 (Advance Electronics Limited; Wrexham, UK) in order to record the differences in output in response to the different stimuli. The stimuli tested were the Face (Starer Absent) and Face + Remote Stare (Starer Absent) conditions from Experiment 3. These two conditions had the same program code except that in the former the code instructed the camera-feed to the starer’s monitor to be masked, and in the latter condition it was unmasked. This code was the same regardless of whether or not the starer was physically present (as per the experimental manipulation of the third experiment). The face image displayed on the staree’s screen was the identical file for both conditions (and indeed, for all of the experiments).

**Results and Discussion**

The first test was to examine the different stimuli for any differences in the overall output of the photodiode (and therefore the luminance) for the full 5,000 ms of exposure. There was no difference, with both conditions providing a mean output of 266 mV.

The second test was a more specific analysis examining the luminance profiles at the onset of the image display. The test revealed a small difference between the two conditions, with the image in the Face + Remote Stare (Starer Absent) condition taking slightly longer to step up incrementally to full luminance than the image in the Face (Starer Present) condition. This difference lasted for approximately 20 ms and corresponded with a difference of approximately 2.5 cd/m² (candela per meter squared; approximately 0.2 lux or 0.7 foot-lambert). As revealed by the first test above, this difference did not continue beyond the first 20 ms as the screen was ramping up to full luminance.
These results suggest that, although there was no difference between the luminance levels of the images on the staree’s screen once the image reached its full luminance level, there was a small difference between the images when they were being initially presented on the screen. This may have, in turn, had an impact upon the corresponding electrocortical processing of that image or provided the participants with some information concerning the particular condition they were experiencing at any one time.

**Overall Discussion**

The body of research presented here initially began as an exploration of the potential electrocortical activity associated with the processing of remote staring detection. However, as it progressed it became an investigation of a possible artifact that has the potential to impact upon a wide range of cognitive neuroscience and psychophysics studies, particularly those that employ event-related measures of electrical brain activity. The most parsimonious explanation for the effects reported in this paper is that they represent the ability of the human brain to process very small and rapid luminance differences between visual stimuli. The mere possibility of the luminance effect providing condition-relevant information that could be processed by the participant potentially undermines any claims of a remote staring detection effect. It should be noted that this potential artifact is related specifically to the methods utilised in this paper, specifically exposing the staree to conventional stimuli concurrently to a remote stare. This would not apply to previous remote staring detection studies, as they did not use this methodology and they also employed the comparatively slow measure of skin conductance. However, whilst it is true that the findings of the third experiment and the photodiode study do suggest a possible luminance difference between the two conditions, the differences involved are so small, namely 2.5 cd/m$^2$ for 20 ms, that they represent a potential anomaly in their own right. Previous research has not found significant changes in electrocortical processing for such small luminance differences.

There is relatively little research in the psychophysics literature exploring the luminance detection threshold in isolation. This is at least partially because it is a difficult phenomenon to test and is reliant upon a multitude of other environmental and psychological factors. The absolute threshold of human luminance difference detection is approximately 0.00001 cd/m$^2$ after 40 min in absolute darkness (Kolb, Fernández, & Nelson, 2012), but as participants in the experiments reported here were not in absolute darkness it is an inaccurate benchmark. It has been noted that luminance differences can be perceived as low as 0.75 cd/m$^2$ (Peli, Yang, Goldstein, & Reevesi, 1991), or even as low as 0.005 cd/m$^2$ (Plainis & Murray, 2000), but the authors in both studies note that due to the perception of luminance differences being logarithmic (i.e., the Weber-Fechner law) and due to the artificial nature of these psychophysics studies, the reliability of these values is to be questioned. One of the main issues is that the majority of psychophysics studies use threshold detection, which involves conscious awareness. However, electrocortical processing studies such as those reported here do not necessarily involve conscious processing, and there are few studies that examine subliminal processing of luminance shifts. In fact, a study that used ERPs to examine the processing of luminance differences failed to find significant effects when comparing the processing of bright (15.5 foot-lambert) versus dim (0.4 foot-lambert) stimuli (Johannes, Münte, Heinze, & Mangun, 1995). As Johannes et al. failed to find significant differences in the processing of simple stimuli with over a 15 foot-lambert difference between them, it is problematic that stimuli with approximately only a 0.7 foot-lambert difference between them for only 20 ms at stimulus onset found in this study could have such a significant impact upon global electrocortical processing. Additionally, a pure luminance processing effect is an elementary feature of a stimulus and should theoretically have an effect on only relatively early components and not on faces that are processed comparatively late (see Allison et al., 1999).

Therefore the findings presented here offer an interesting problem to parapsychologists, psychophysicists, and cognitive neuroscientists. The problem for parapsychology is that ideally these experiments need to be repeated—controlling for any potential luminance differences—in case these results do not represent an artifact and remote staring detection does have an impact upon global electrocortical processing of other stimuli. For psychophysics, these results may demonstrate the complexity of basic luminance differences and the possibility of examining these using “below-threshold” detection methods. Finally, for cognitive neuroscience, these results open the possibility that the often-published effects noted in basic face and object processing may be confounded by very small differences in the luminance levels of different stimuli.


Anomaly of an Anomaly: The Cortical Electrophysiology of Remote Staring Detection


Rosenthal, G. T., Soper, B., & Tabony, R. S. (1994). *Student beliefs concerning the ability to detect when someone is watching: A revision of the Paranormal Belief Scale*. Unpublished manuscript, Nicholls State University, Thibodaux, LA.


Acknowledgments

The research reported here represents part of the first author’s PhD thesis (Baker, 2007) conducted at the University of Edinburgh. We would like to dedicate this paper to the memory of Professor Robert Morris, who was Ian Baker’s primary PhD supervisor until his untimely death. We would also like to thank...
the INOVA foundation, the Bial Foundation, and the Society for Psychical Research for their financial support. We would also like to thank Paul Staples and Professor David Sheffield for their comments on earlier drafts of this paper.

Abstracts in Other Languages

French

UNE ANOMALIE D’ANOMALIE : ETUDE DE L’ELECTROPHYSIOLOGIE CORTICALE DE LA DETECTION DU REGARD À DISTANCE

RESUME : S’il y a des éléments de preuve d’un effet global de la détection du regard à distance, alors il devrait théoriquement y avoir d’autres preuves d’un processus électrophysiologique de traitement de cette information dans le cerveau. Une série de 3 expérimentations examinant les corrélats électrocorticaux potentiels de la détection du regard à distance est présentée, suivie par une 4e expérience pour examiner un artefact potentiel. La 1ère expérience fournit une exploration initiale de cet effet, montrant d’abord que « la détection du regard à distance » ne correspond pas de façon évidente à un processus repérable dans le temps mais agit plutôt sur les processus se produisant au même moment. La 2e expérience fournit des preuves que cet effet n’est pas spécifiquement en lien avec le processus de reconnaissance de visage mais peut avoir un impact sur d’autres processus. La 3e expérience met en lumière un potentiel artefact qui pourrait expliquer l’effet de « regard à distance », qui est vérifié dans l’expérimentation finale. Les résultats globaux sont discutés à la lumière d’un subtil et intéressant effet de luminance psychophysique qui pourrait potentiellement avoir un impact sur une large variété d’expérimentations qui emploient des mesures relatives à des événements de processus électrocorticaux.

Spanish

ANOMALÍA DE UNA ANOMALÍA: INVESTIGANDO LA ELECTROFISIOLOGÍA CORTICAL DE LA DETECCIÓN DE SER OBSERVADO A DISTANCIA

RESUMEN: Si hay evidencia de un efecto general de poder detectar si alguien nos observa a distancia, teóricamente también debería haber evidencia del procesamiento electrofisiológico de dicha información en el cerebro. Presentamos una serie de 3 experimentos que examinaron posibles correlatos electrocorticales de la detección de ser observado a distancia (DOD), seguidos por un cuarto experimento para examinar un posible artefacto. El primer experimento fue una exploración inicial de este efecto y encontró principalmente que DOD no está asociado a un procesamiento sincronizado evidente, sino que actúa sobre otros procesos que ocurren al mismo tiempo. El segundo experimento proporcionó evidencia de que este efecto no está relacionado específicamente con el procesamiento de rostros, pero puede tener un impacto en otras formas de procesamiento. El tercer experimento mostró evidencia de un posible artefacto que podría explicar el efecto DOD, verificado en el experimento final. Discuto los resultados globales a la luz de un interesante y sutil efecto psicofísico de luminancia que podría tener un impacto en una amplia variedad de experimentos que emplean medidas relacionadas con los eventos de procesamiento electrocortical.

German

EINE ANOMALIE EINER ANOMALIE: ZUR UNTERSUCHUNG DER KORTIKALEN ELEKTROPHYSIOLOGIE BEIM NACHWEIS DES BEOBACHTETWERDENS (REMOTE STARING)

Verarbeitungsprozess darstellt, sondern sich vielmehr auf andere Prozesse auswirkt, die gleichzeitig ablaufen. Das zweite Experiment ergab Hinweise darauf, dass sich dieser Effekt nicht spezifisch auf die Gesichterverarbeitung auswirkt, sondern auch andere Verarbeitungsformen beeinflussen kann. Das dritte Experiment fand einen Hinweis auf ein mögliches Artefakt, das den „remote staring-Effekt“ erklären konnte, was im letzten Experiment bestätigt wurde. Die Gesamtergebnisse werden im Lichte eines interessanten und subtilen psychophysikalischen Luminanzeffektes diskutiert, der möglicherweise für eine größere Anzahl von Experimenten von Bedeutung sein könnte, die ereignisbezogene Messungen der elektrokortikalen Verarbeitung verwenden.
ANOMALOUS REMOTE DIAGNOSIS: MENTAL AND MOTOR
PSI IMPRESSIONS UNDER ICONIC REPRESENTATION
OF THE PERSON-TARGET

BY ALEJANDRO PARRA AND JUAN CARLOS ARGIBAY

ABSTRACT. A common procedure to encourage self-claimed psychics is to use iconic representations (photographs) as inductors. The aim of this study was to compare 2 conditions – “mental” and “motor” – using images of the faces of persons as targets. Specifically, we wanted to determine if the scores were different based on 2 kinds of stimulus (diseased and healthy). Participants were clustered as psychics and nonpsychics. The sample consisted of 224 participants, recruited through an e-mail list, and their ages ranged from 18 to 75 (M = 44.19 years old). Eight photographs included 4 of subjects “sick” with a medically diagnosed disease and 4 of healthy persons (the “controls”). All of the images of sick persons were taken from 6 months to 2 years before the diseases were discovered. The results of the 2 procedures were significantly above chance: the mean “mental” score = 2.32 (p < .001, 1-tailed), and the mean “motor” score = 2.15 (p = .016, 1-tailed). A second analysis was carried out to determine if the mental and motor procedures led to different scores; indeed, mental scored significantly higher than motor (the mean “mental” score = 2.32 vs. the mean “motor” score = 2.15 (p < .001).

Keywords: psychics, psychometry, remote diagnosis, dowser

Anomalous detection, or psychic diagnosis, is reported frequently by healers and psychics. They may include several components—such as unusual awakenings, understandings based upon sensory observations, and data analysis—which are suddenly perceived in a novel relationship and order, or data based on information from anomalous cognitive processes—such as telepathy or clairvoyance—which appear to transcend ordinary reasoning (Benor, 1992).

Psychics claim that such impressions may sometimes be obtained through scanning the “energy body,” and a number of them have gained reputations as psychic detectives using psychometry, such as Ossowiecki (Stevenson, Barrington, & Weaver, 2005) and Croiset (Anderson, 2006; Pollack, 1964). One of the most impressive examples of this approach was the American psychic healer Edgar Cayce (Stearn, 1967), who was able to provide accurate diagnoses given only the name and address of participants, some of whom were many miles away. However, there are few reports on quantitatively evaluated studies with psychics. Vaughan (1974), Shealy (1988), Young and Aung (1997), Mison (1968), and Brier, Savits, and Schneiderler (1974) reported results indicating some correspondences between the psychic diagnoses and medical records of patients, but these correspondences tended to be insufficiently impressive to warrant considering psychic diagnosis as a useful alternate method for diagnosing disease.

A common procedure for self-claimed psychics to obtain extrasensory or psi impressions is to use iconic representations (photographs) as inductors (Rogo, 1974). The use of an inductor is by no means a requirement, and many psychics provide impressions about target persons without involving any object, using instead, for instance, just the name and age of the target person. Another procedure, dowsing—the supposed evocation of unconscious responses by use of ideomotor responses magnified by a physical device, such as forked twigs or a pendulum—has long been espoused as a technique for helping individuals to utilize untrained psi abilities (e.g., Eastwood, 1993). Laboratory research on the direct mental influence of living systems (DMILS) and the detection of remote staring suggests that an individual’s conscious response may not be a good measure of psi, whereas such an individual may indeed show a physiological reaction to certain psi-mediated stimuli (e.g., Sah & Delanoy, 1994). If this is the case, then a system looking at the physiological responses of individuals undertaking a psi task might be useful in helping them to recognize when they were using psi. In effect, it would be a technological version of the old dowsing devices.

This ideomotor phenomenon, repeatedly described in the medical literature over 150 years ago by William Carpenter (1852) and later elaborated upon by the noted psychologist William James (1890), is clearly not a recent discovery. A very good overview is presented by Herman Spitz in his book Nonconscious...
The interesting question in relation to the pendulum’s use in dowsing is whether these unconscious muscular movements may be produced by extrasensory perception, or psi. Normally, dowsers use the pendulum to get answers to questions that they are interested in. Dowsers (and healers as well) will need to decide in advance what each direction of movement means. The pendulum simply acts as an amplifier of subconscious ideomotor movements, so that they can be more easily detected.

An important question to consider is whether the motor system of the subject is influenced indirectly, translating certain “mental” impressions to a “motor” form, or directly, if the subject has no intervening “mental” impressions but only “motor” ones. Or, stated in a more general way, the question is whether a percipient is better able to locate the target paranormally when asked to imagine it (the “mental” way of receiving) or when asked merely to designate it directly with a simple motor response (the “motor” way of receiving). For instance, healers report intuitive awareness and understanding of patients’ physical, emotional, mental, and spiritual problems (Benor, 2002; Krieger, 1979) as words appearing in their mind, including technical diagnoses that they themselves do not comprehend but which doctors later confirm to be accurate (Stearn, 1967); as body sensations—especially pain—that reflect those of patients; as smells; as visual images of organ dysfunctions; or simply as an “inner knowing” (Brennan, 1987; Freed, 1992; Schwartz, 1967).

Both psychics and their clients must be sincerely impressed on occasion by the achievements that confirm the psychics’ belief in their abilities, mainly in terms of face-to-face interventions. The face is often an important source of identification with others and conveys significant social information (Nelson, 2001; Bruce & Young, 1986). It is known that early perceptual experience is crucial to the development of visual perception; this orienting response undoubtedly encourages the rapid development of face-specific skills such as the ability to identify friendly others and relatively complex preverbal communication. There are seven distinct types of information that we derive from seen faces, namely: pictorial, structural, visually-derived semantic, identity-specific semantic, name, expression, and facial speech code. Whatever the source of psychics’ statements, their experience and expertise in certain areas might make it worth consulting them for advice, even if they were to make no genuine paranormal statements in a given session.

A series of psychometry-based experimental sessions was designed. We explored the “token-object” effect in groups in the context of a program on psi development based on Tart’s learning theory of ESP (1975). Our design here is similar to a previous one (Parra & Argibay, 2007b), where we used personal objects (comb, handkerchief, hair brooch, or billfold) from four adult volunteers who suffered from medically diagnosed diseases. Participants, who were clustered as psychics and nonpsychics, performed trials of psychic diagnosis. Although neither group obtained highly significant results, we found high variability in a positive direction for the psychics and in a negative direction for the nonpsychics ($p < .05$). It appears that psychic diagnosis relates to perceptions of “information” in and around the person-target, and that these may be difficult to translate into physical diagnoses.

Following on a number of previous experiments (Parra & Argibay, 2007a, 2007b, 2008, 2009a, 2009b), ordinary people (nonpsychics) were tested using photographs of the faces of sick persons as targets. We wanted to explore two strategies for using and appraising the so-called token-object effect. Instead of objects, photographs of the faces were used along with the nicknames of the person-targets. The aim was to compare two conditions—“mental” and “motor”—using images of the faces of sick persons as targets. Specifically, we wanted to (a) determine if the participants scored differently with the two kinds of stimulus (diseased and healthy persons), and (b) determine if the whole sample of participants used both kinds of procedures (mental and motor).

**Method**

**Participants**

The sample consisted of 224 participants, 208 (76%) females and 68 (24%) males, all of whom were well-educated and believed in psi. Their ages ranged from 18 to 75 ($M = 44.19, SD = 12.90$). Personal experiences suggestive of psi were reported by the majority of the participants, such as having experienced
ESP feelings around sick people (58%), around past place events (50.8%), around token-objects (34.7%),
and around token-photos (38.3%). Participants were recruited through an e-mail list, and an announcement
was also placed on a web page (www.alipsi.com.ar). The announcement provided a brief explanation of
the test procedure and encouraged people to schedule an interview with us in order to obtain more
information.

**Participant Setting**

Seventeen separate groups were tested by the first author (AP) and the second author (JCA) at the
IPP headquarters in Buenos Aires, in 2-hr sessions over a period of 3 years. There were between 5 and 10
participants in each group. AP and JCA aimed at creating a friendly and informal social atmosphere.

Two rooms were required to conduct the experiment: one for AP and the participants, and one for
JCA. The participants were seated in chairs and tested in groups. All participants were present together
when handling the photographs, but they operated individually; no interaction was allowed. AP handed
out envelopes containing the pairs of photographs. Each pair was supplied with an answer sheet including
written test instructions (although instructions were also given verbally). The participants were informed
that we were doing a test of ESP using material said to stimulate extrasensory abilities in people.

**Target Material**

Two coexperimenters (RM and JMC), not present during the sessions with the participants, gave
us eight photographs. People were selected as person-targets because they suffered highly symptomatic,
noncontagious diseases. Four of them were “sick” with a medically diagnosed disease at the time of the
experimental session; these photographs were matched with four photographs of healthy persons (the
“controls”). The people featured in those photographs were unknown to JCA and AP. We decided to
use people who had been diagnosed with the following four diseases: Diabetes mellitus (Type 2; Fabiana
S.), Non-Hodgkin’s lymphoma (Héctor M.), Osteoarthritis (Pamela R.), and Anosmia (Romina A.). The
person-targets and conditions are presented in Table 1.

**Security Measures**

All the images of sick persons were taken months or years before the diseases were discovered, so
there were no visual cues of disease in the face. Before each session, the coexperimenters JMC and RM
coded the pairs of photographs of diseased and healthy people in a manner unknown to JCA and AP.
Then they delivered the eight photographs in an envelope to JCA. To avoid direct contact with the original
colour photos, and to preserve their print style, JCA scanned the photographs in black and white and then
printed them on high-quality glossy paper; he also recoded the pairs of photographs in a manner unknown
to AP. The nicknames were printed on each photograph.

<table>
<thead>
<tr>
<th>Nickname</th>
<th>Disease</th>
<th>Test condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Héctor M.</td>
<td>Non-Hodgkin’s lymphoma</td>
<td>motor</td>
</tr>
<tr>
<td>Roxana M.</td>
<td>Healthy (“control”)</td>
<td>motor</td>
</tr>
<tr>
<td>Fabiana S.</td>
<td>Diabetes mellitus</td>
<td>mental</td>
</tr>
<tr>
<td>Andrea A.</td>
<td>Healthy (“control”)</td>
<td>mental</td>
</tr>
<tr>
<td>Romina A.</td>
<td>Anosmia</td>
<td>motor</td>
</tr>
<tr>
<td>Laura S.</td>
<td>Healthy (“control”)</td>
<td>motor</td>
</tr>
<tr>
<td>Pamela R.</td>
<td>Osteoarthritis</td>
<td>mental</td>
</tr>
<tr>
<td>Liliana P.</td>
<td>Healthy (“control”)</td>
<td>mental</td>
</tr>
</tbody>
</table>
Using a list of random numbers, JCA selected the order in which the photographs (sick/healthy persons) were used for the tests. JCA put two photos of sick persons and another two photos of healthy persons, in a counterbalanced way, in an envelope to be used for both conditions, mental and motor (Ntrials = 8). All procedures related to the presentation of the targets were randomized using an RNG.

JCA did not enter the room during the testing but remained in a nonadjacent, sound-attenuating room. The details of all these procedures were unknown to AP, so that he remained unaware of which photographs corresponded to the diseased/healthy conditions. Once the experimental sessions had been completed for the whole group, AP handed the photographs back to JCA, who recoded them to their original state and then returned them. This procedure was repeated for each group.

Consent Form

Participants signed an appropriate consent form, in easily comprehensible language. The form specified that the person (a) had the capacity to consent, (b) had received all significant information about the procedure, (c) had freely and without undue influence expressed consent, and (d) was aware that the consent had been appropriately documented (cf. Beahrs & Gutheil, 2001). Joining the group was voluntary, and all data collected were treated confidentially.

Instructions

Instructions on each form were related to short descriptions about the sickness symptoms in easy, nontechnical language:

1. **Diabetes mellitus**: It often involves frequent urination and increased thirst, with a consequent increase in fluid intake. There may also be weight loss (despite normal or increased eating), increased appetite, and irreducible fatigue. Thirst develops because of osmotic effects, that is, overly high glucose in the blood is excreted by the kidneys. Another common symptom is altered vision. Especially dangerous symptoms include the smell of acetone on the patient’s breath (a rapid, deep breathing), and any altered state of consciousness or arousal.

2. **Non-Hodgkin’s lymphoma**: This is a type of cancer that originates in the lymphatic system. The NHLs are a diverse group of blood cancers that include any kind of lymphoma except for Hodgkin’s. They vary significantly in their severity, from marginal to very aggressive. Symptoms and signs of non-Hodgkin’s lymphoma include swollen lymph nodes and fever.

3. **Knee arthritis**: This is usually a slowly progressive degenerative disease in which the joint cartilage gradually wears away. It most often affects middle-aged and older people. Pain associated with arthritis develops gradually. The joint may become stiff and swollen, making it difficult to bend or straighten the knee. Pain and swelling are worse in the morning or after a period of inactivity. Pain may also increase after activities such as walking, stair climbing, or kneeling. The pain may often cause a feeling of weakness in the knee, resulting in a “locking” or “buckling.”

4. **Anosmia**: Lack of functioning olfaction, or, in other words, an inability to perceive odors. It is caused by chronic meningitis that increases intracranial pressure over a long period of time. People experience sudden-onset anosmia, finding food less appetizing, though congenital anosmics rarely complain about this. The loss of smell can also be dangerous because it hinders the detection of gas leaks, fires, and spoiled food.

Task Conditions

The experimenters delivered two photographs in little envelopes to each participant. For each pair of photographs (diseased/healthy persons), a form contained printed instructions for the participants (in addition to instructions given verbally). Before completing the ESP test, each participant underwent a 9-min relaxation exercise for both conditions. The experimenters told the participants that they were doing an ESP test, adding that the test could stimulate psychic diagnosis abilities in people. AP remained in the room as a silent observer throughout the testing period, which lasted about 40 min. The participants did not provide any verbal reports; instead, they just checked “diseased” or “healthy” for each photograph.
Once the participants had completed the answer sheets for each pair of trials, they passed the envelopes to AP, who handed them and the answer sheets back to JCA for recoding. This procedure was repeated for each participant. Participants were not given any trial-by-trial target feedback during the testing period. The trials were performed in a counterbalanced way, four using the “mental” procedure and four using the “motor” procedure.

1. **Mental condition.** Experimenters asked the participants to “remain with eyes closed, quiet, waiting for mentations about the object during a few minutes.” Participants remained with their hands over the photograph stimulus, waiting to receive impressions. The aim of the mental procedure was to obtain a combination of impressions, feelings, intuitions, and imagery related to the photograph-target. Afterwards, participants wrote the nicknames printed on the photographs on a form.

2. **Motor condition.** Experimenters asked the participants to “remain with the pendulum in their hand (right or left), quiet, waiting for responses magnified by a physical device, namely, the pendulum.” Participants held the pendulum between the thumb and first finger of the dominant hand, with about 7 in. of string to the point of the pendulum (a small ball of wood), which was about 1/4 inch (0.5 cm) above a stimulus photograph, waiting to receive impressions for movement of the pendulum. The yes–no responses were obtained by coding clockwise or counter-clockwise pendulum movements. Afterwards, participants wrote the nicknames printed on the photographs on a form.

### Results

The experiment examined the performance of two conditions (mental and motor) using a procedure featuring photographs of diseased and healthy persons in a forced-choice test. Four trials were performed for “mental” (psychometry) condition and four trials for “motor” (pendulum) condition, so the \(MCE\) is \(4 \times .5 = 2\) per each condition (diseased/healthy; see Table 2).

#### Table 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>(N)</th>
<th>(TP)</th>
<th>(MCE)</th>
<th>(M)</th>
<th>(SD)</th>
<th>One-sample (t)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental</td>
<td>224</td>
<td>4</td>
<td>2</td>
<td>2.32</td>
<td>1.00</td>
<td>4.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Motor</td>
<td>224</td>
<td>4</td>
<td>2</td>
<td>2.15</td>
<td>1.04</td>
<td>2.17</td>
<td>.016</td>
</tr>
</tbody>
</table>

*Note.* TP means target persons.

As shown in Table 2, the two groups scored significantly above chance: Mean “mental” = 2.32, \(t(223) = 4.83; p < .001\); and Mean “motor” = 2.15, \(t(223) = 2.17, p = .016\).

#### Table 3

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mental</th>
<th>Motor</th>
<th>Paired-samples (t)</th>
<th>(p)</th>
<th>Effect size ((r^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M (SD))</td>
<td>2.32 (1.10)</td>
<td>2.15 (1.04)</td>
<td>3.61</td>
<td>&lt;.001</td>
<td>.055</td>
</tr>
</tbody>
</table>

A second analysis was carried out to determine if participants using the mental and motor conditions scored differently with the two kinds of stimulus (diseased and healthy persons). As shown in Table 3, “mental” scored significantly higher than “motor” condition: “Mental” = 2.32 versus “Motor” = 2.15, \(t(223) = 3.61, p < .001\).
In the present study, we conducted a number of trials of psychometry with a group of ordinary people. The aim was to determine if the whole sample of participants scored differently with the two kinds of procedures (mental and motor conditions) and if they scored differently when reacting to photographs of sick versus healthy persons. The two groups scored significantly above chance. We conclude that this experiment offers support for the claim that iconic representation through mental condition (psychometry)—implying mental (visual) representation of the person-target—is psi conducive. In this case we found a significant difference between the mental and motor conditions, notably in a positive direction for the psi impressions/imagery. The iconic representation appeared to favour psi because it was more ecologically valid.

Also, a substantial number of participants indicated having had some training in meditation or other techniques involving psychic abilities and/or internal focus of attention; thus maybe psychometry (the mental procedure) favoured psi itself. However, some problems were involved in terms of participants’ impressions, as some of them had difficulty expressing their feelings and sensations or describing the target person (through imagery), or experienced psychological resistance (fear of psi) or trouble using the pendulum (mainly in people untrained in dowsing). Future experiments should explore a nonvisual approach (with photographs or objects), under both conditions, in order to test a “blind” condition using a pendulum. Psi seemed to work better in the mental than the nonmental condition, such as using visual targets instead of token-objects as stimulus (Parra & Argibay, 1997a). It may be that a mental condition is more congruent with the way ESP functions. Finally, a mental condition may add motivation compared to the pendulum condition, and that could also facilitate the psi task.

We are working under the assumption that psi results from the detection of a weak energetic signal propagating in space and/or time. Psychometry continues to be an area for exploration into ESP, and there is a logical basis for further experimentation in this area. However, at this time it may require the services not only of parapsychologists but also of—particularly—biologists, neurologists, and physicists. Certainly a biophysiological approach to the issue of ESP seems called for. Moreover, it is proposed that a psi signal contains second-order information about the psi source’s physiological reaction to the stimulus. This signal might be affected by poor encoding (i.e., the cognitive functions of the transmitter), as well as noise (an environmental factor) and psychological filters applied by the receiver. At the same time, the information transmitted may be second-order, that is, relating to how the transmitter’s brain reacts upon perceiving the target stimulus. In other words, when perceiving the original stimulus, the transmitter’s brain produces a specific pattern of activity. Further research should examine that possibility.

References


*Salta 2015 (C1137ACQ)*

*Acknowledgments*

The authors would like to thank the Bial Foundation for financial support of this research project. Thanks also for the generous collaboration of Juan Manuel Corbetta (JMC) and Romina Mielgo (RM), who acted as coexperimenters and provided us with photographs of relatives and friends.
DIAGNOSTIC ANOMAL A DISTANCE : IMPRESSIONS PSI MENTALES ET MOTRICES PORTANT SUR LA REPRESENTATION ICONIQUE DE LA PERSONNE-CIBLE

RESUME: Une procédure commune pour encourager les sujets psi auto-proclamés est d’utiliser des représentations iconiques (photographies) en tant qu’inducteurs. Le but de cette étude est de comparer deux conditions – « mentales » et « motrices » – en utilisant des images des visages des personnes en tant que cibles. En particulier, nous avons cherché à déterminer si les scores étaient différents en fonction des deux types de stimulus (malade ou sain). Les participants étaient regroupés en « sujets psi » et « non-sujets psi ». L’échantillon comportait 224 participants, recrutés sur une liste électronique, et leurs âges allaient de 18 à 75 ans (M = 44,19 ans). Le matériel était composé de huit photographies incluant quatre de personnes « malades » (avec une maladie diagnostiquée médicalement) et quatre de personnes saines (les « contrôles »). Toutes les images des personnes malades furent prises de 6 mois à 2 ans avant que la maladie ne soit découverte. Les résultats des deux procédures furent significativement supérieurs au hasard : le score « mental » moyen = 2.32 (p < .001, one-tailed), et le score « moteur » moyen = 2.15 (p = .06, one-tailed). Une seconde analyse fut réalisée pour déterminer si les procédures mentales et motrices menaient à des scores différents ; en effet, le score « mental » moyen était significativement supérieur au score « moteur » moyen (p < .001).

Spanish

DIAGNOSTICO ANOMALO A DISTANCIA: IMPRESIONES PSI MENTALES Y MOTORAS BAJO LA REPRESENTACION ICONICA DE LA PERSONA OBJETIVO

RESUMEN: Un procedimiento común como estimulo a individuos que dicen ser psiquicos es el uso de representaciones iconicas (fotografias) como inductores psi. El objetivo de este estudio fue comparar dos condiciones – “mental” y “motor” – usando imgenes de los rostros de personas como objetivos. Específicamente, pretendemos determinar si los resultados fueron diferenciales en base a dos tipos de estímulos (personas enfermas y no enfermas). Los participantes se re-agruparon como psiquicos y no psiquicos. La muestra consistio de 224 participantes, reclutados a través de e-mails. Sus edades oscilaban entre 18 a 75 (M = 44.9 años de edad). Ocho fotografias mostraron cuatro individuos “enfermos” con una enfermedad medicamente diagnosticada y cuatro de personas sanas (“controles”). Las imágenes de personas enfermas fueron tomadas de 6 meses a dos años antes que las enfermedades fueron descubiertas. Los resultados de ambos procedimientos dieron significativamente por encima del azar: la puntuación media de “mentales” = 2.32 (p < .001, a una cola), y la puntuacion media de “motoras” = 2.15 (p = .06, a una cola). Se realizo un segundo análisis para determinar si el procedimiento mental o motora produjeron resultados diferenciales, lo cual, de hecho, el procedimiento mental obtuvo puntuaciones mas altas que la motora (“mental” = 2.32 vs. “motor” media = 2.15; p < .001).

German

ANOMALE FERNDIAGNOSE: MENTALE UND MOTORISCHE PSI-EINDRÜCKE BEI BILDLICHER DARSTELLUNG DER ZIELPERSON

OBITUARY

INGO DOUGLAS SWANN
1933–2013

BY EDWIN C. MAY

Ingo Swann, born September 14, 1933, in Telluride, Colorado, passed away on February 1, 2013, in New York, New York.

Ingo’s contribution to parapsychology is incalculable. His early activity in what eventually became known as remote viewing was in 1971–1973 when he participated in out-of-body trials at the American Society for Psychical Research in New York. In these experiments, Ingo was asked to describe the content of a sealed box a meter or so above his head.

His early work, in what is now known as Star Gate—the U.S. government’s 20-year, 20-million-dollar program to study and use ESP as a collection asset for intelligence during the Cold War, is well documented in *Mind-Reach* (Targ & Puthoff, 1977). Rather than enumerating all of Ingo’s significant contributions to the field, I thought I would tell a personal narrative to illustrate this remarkable and complex individual.

I first met Ingo at a party in his flat near 3rd Street in Manhattan, during which Ingo learned of my own interest in psi and that I was halfway through a lengthy research adventure in India. Later we met for dinner in the Village, where Ingo told of his work in California in rather vague terms. I did not know it at the time, but he was unable to provide specifics given that the ESP program at Stanford Research Institute (SRI) at the time was highly classified.

I returned from India in the spring of 1975 and began working at Maimonides Medical Center with Chuck Honorton, mainly on RNG studies. Ingo was quick to volunteer for some of those studies. He was always a willing and intently curious participant. In particular, he was fascinated with the random number generator (RNG) work coupled with the idea of separately counting trials and hits when he was in a particular EEG alpha bandwidth and power. We reported this work at an American Physical Society meeting that year. Ingo expressed great interest in that I was an experimental physicist as opposed to a theoretical one.

I later found out that Ingo encouraged Hal Puthoff to hire me at SRI as consultant (and later as a senior research physicist) to help with an ongoing, Navy-funded PK experiment with a magnetometer, and with ongoing EEG studies. Ingo told me then that he did not want some critic to invalidate his PK work because someone threw a light switch down the hall. Thus, I owe my job to Ingo.

Now that we were both working with Hal Puthoff and Russell Targ at SRI, Ingo and I grew close as friends. He began to introduce me to a variety of his art friends, which included a trip to Mexico to restore some artwork there that had been damaged by a hurricane. I was fortunate to watch Ingo work on some of his best known paintings, some of which are now hanging in the executive section of the Smithsonian’s Air and Space Museum. *Aft Ship’s View of Sagittarius* is a 20-foot masterpiece and was my first introduction to Ingo’s ability for intensive concentration and dedication to work. Always puffing away on his cigars, he might work for many hours uninterrupted. Ingo’s flat did not have a proper kitchen, but he had a large number of “hot-plate” specials that bordered on the gourmet. And, of course, there was always the hamburger joint next door.

From my perspective, Ingo’s major contribution to psi was with his training methodology. During the late 70s and early 80s Ingo’s dedication and work ethic were far superior to any of us in the field! He was studying conditioning theory and, in particular, operant conditioning, which took him to the Stanford University library for sometimes as long as 12 hours. What he gleaned from all that work was then and remains today his primary contribution to the field.

At SRI in those days, we all had noticed that the data early on in a remote viewing session tended to be more accurate than the later data. What Ingo did was to formalize this in his special variant of operant
conditioning. Psychologists have often used word association tests in therapy to try to get at what is going on emotionally with their clients. Say the first word that comes to mind when you hear, for example, “fast.” The idea here is that the client would not have enough time to fabricate a response in order to “please” the therapist.

So Ingo invented a stimulus/response methodology to incorporate this notion into remote viewing sessions. The stimulus was not a word but rather the geographical coordinates of a randomly selected site on the globe. An RV trainee was asked to respond rapidly but within a tight structure. In fact, Ingo was so adamant about structure, he posted signs in the laboratory that read: “Content be damned; structure is all that matters!”

This is not the place to provide a scientific critique of his overall training procedure that was eventually adopted by the newly established Army Ft. Meade remote viewing group. It suffices to say that the stimulus/response aspect of his training method was a jewel of wisdom, but many of the other aspects, while seeming straightforward within the context of classical conditioning theory, were not especially helpful.

Ingo had, at times, a wicked sense of humor, and if you were going up against him you might be the victim of one of his intense gazes that could melt lead at 100 meters! One time we were giving a joint presentation to a group of secondary education teachers in the Bay Area of San Francisco. During the Q & A portion, one teacher asked the rather hostile question, “Since you are a Scientologist who believes in reincarnation, why not get rid of your fat, 40ish body and come back in a nicer one.” Ingo responded, “Well, what keeps me in this fat, 40ish body is the threat of secondary education all over again!” That brought down the house.

Ingo loved to push the social envelope by deliberately being outrageous. We both appreciated ballet, and after seeing the movie *Turning Point*, Ingo decided that nothing would do but he must take ballet lessons at a studio near SRI. So on the days of lessons, he would change into his tights on the third floor of our building at SRI and nearly prance his way through the multitude of engineers at their desks on the first floor for dramatic effect. On one of these occasions, Ingo insisted that Hal, Beverly Humphrey, and I come to watch his lesson. We were surprised that the next youngest person in his class was probably 7 or 8 and limber as a piece of string. But Ingo was struggling in the warm-up to even touch his feet while sitting on the floor! Ingo was not at all amused given that the three of us could hardly contain our laughter. The good news was that Ingo took it all in and continued his lessons for some time.

A book could easily be written about Ingo and his many idiosyncrasies—some funny, some not—but always engaging and interesting.

Ingo will be remembered for his contribution to psi research, and some day when the data are released, his contribution to gathering intelligence. The world has yet to learn of this remarkable man’s contributions in that arena.

Ingo’s last activity in Star Gate was in 1986, when he participated in an extensive magnetoencephalograph study we did at Los Alamos National Laboratory. He told me near the end of that work that he was tired of research and would rather focus upon his art, writing, and other things. We parted then as colleagues and friends.

Travel well, dear Ingo. You are missed.

Reference


*Laboratories for Fundamental Research*

330 Cowper St., Suite 300
Palo Alto, CA 94301, USA
may@lfr.org
BOOK REVIEWS


Exploring Frontiers of the Mind-Brain Relationship is an important contribution. It provides a balance of the purely physical analyses of consciousness with the psychobiological. In this way, it attempts to bridge mind philosophy with science. It empirically examines transcendent phenomena, such as mystical altered states, near-death experiences, end-of-life, mediumistic experiences, and past-life memories. Consequently, this book’s 12 chapters confront the hard scientific dilemmas, and its cast of scientists should be commended. The book has a good mixture of 15 often well-known scientists and physicians from six countries, some prominent in the consciousness area, others solidly based in brain functioning, philosophy, history, and physics. It reflects the accumulated data of dozens of careful researchers over more than a century.

The book relies heavily, as it should, on the empirical findings of the neurosciences but also includes philosophical perspectives. It has four major sections:

1. The first is “Philosophy and History,” where the book moves away from materialistic reductionism, arguing for alternative approaches. The first two chapters discuss the challenges of the mind-brain problems (Saulo de Freitas Araujo and then Robert Almeder). These chapters not surprisingly do not cover unified monism (which is so new its omission can be understood), but I think it could have covered appropriately panpsychism (mentioned later in the “Physics” section) and neutral monism. The third chapter in this section is “History,” and Carlos S. Alvarado, as always, does an excellent job.

2. “Physics” is the next section. It reflects quantum brain biology perspectives and sees consciousness not as it is generally regarded, as an epiphenomenon of the brain. Chris Clarke links “No-collapse Physics with Consciousness,” and Stuart Hameroff and Deepak Chopra discuss the “Quantum Soul.” These are both topical and important areas. I would have liked to see such controversial areas as the Copenhagen interpretation of physics, entanglement, and the Aspect experiments and their follow-up discussed in detail. I think these would have been pertinent.

3. “Functional Neuroimaging,” reflects neurophysiology. This is discussed in some detail, providing insights into the intricacies of the human brain as well as some of its limitations, including the brain as a computer model as well as functional neuroimaging, where scanning, plus much of the content and function of mind can be related to brain location or function. I think this is a narrow section, and I would have preferred it to cover specifically neuroimaging. Specifically discussed are the correlates of “Meditation and Mindfulness” (Jesse Edwards, Julio Monti, and Andrew B. Newberg), which appears to be significantly incomplete, and “Functional Neuroimaging” itself by Mario Beauregard. This section should cover far more than specifically neuroimaging. The Persinger temporal lobe work is cited, but inexplicably the Neppe (1983) and Palmer and Neppe (2003, 2004) work is not. I think this whole section on neuroscience is the weakest part of the book, because though much is cited, many key areas such as subjective paranormal experiences and phenomenology, and the frontal lobe work, appear to have been neglected. I would have expected to read more about the presentiment areas, including the contributions of the Standish Bastyr group (e.g., Richards et al., 2005), as well as Bierman and Radin’s ideas (Radin, 2006).

4. “Human Experiences as Promising Lines of Investigation of Mind-Brain Relationship” is the fourth major section: It is a very variable section and possibly could just as easily been called “Survival.” It is generally well done. These discussions allow perspectives for spirituality and for
The Journal of Parapsychology

the mystical, including such phenomena as near-death experiences, out-of-body experiences, and past-life memories, transcendence, and ostensibly nonlocal states. Specifically, “NDEs” (Peter Fenwick) and “Death, End of Life Experiences” (Franklin Santana Santos and Peter Fenwick) are two very interesting discussions, well-informed and with a well-developed balance of history, experience, and theorizing. “Mediumship and the Mind-Brain Relationship” (Alexander Moreira-Almeida) is a chapter I particularly liked, because Brazilian work in the area is so neglected in the English literature and key general dilemmas are raised. “Cases of the Reincarnation Type” follows. Author Erlendur Haraldsson’s perspective is, as always, thorough and knowledgeable, yet it reflects the outer edges of so-called “mind” in this volume.

The very short final 12th chapter by the editors, like the Preface, synthesizes information.

The ideas in the book are broad and range from Buddhist thought to William James’ filter. This book confronts the mind-brain controversies and challenges readers to reformulate their conclusions. It also is balanced, discussing the limitations and strengths of these approaches. Researchers and clinicians across many disciplines can appreciate this book, as its content comprises consciousness research, medicine, neuroscience, neuropsychiatry and behavioral neurology, parapsychology, philosophy, psychiatry, psychical research, psychology (cognitive, personality, and social), and quantum physics. It allows the reader to conceptualize some of the fundamental mysteries of consciousness, and it is this consciousness that has to be regarded as a key theme. The book is not esoteric; it is targeted to more than scientists, so the intelligent, interested layperson may benefit. One important conclusion is that our current materialism is too narrow, and that we can benefit from a top-down approach to consciousness.

The four perspectives are integrated into a broader strategy: In the first two sections, the editors show the philosophical and historical limitations of reductionistic materialism (which is the prevailing perspective for most scientists today) and then present some models in physics that allow for views of mind of nonreductionists. With the next step, in the later two sections, they provide some serious and challenging empirical data in the nonreductionist area. However, the book was not meant to be comprehensive and all-inclusive. Consequently, it did not present many alternative theoretical perspectives for mind, because its purpose was to open a legitimate dent in the mainstream materialistic view, thereby providing inroads for other works advancing alternative perspectives.

So how do I see this book? Clearly it is a significant contribution, but I confess disappointment, not at what it contains but what to me are significant omissions. There is an inadequate perspective of what mind truly is and where it fits into the finite and the infinite. The chapters are a hodge-podge of choices of certain directions, neglecting others. For example, why is the literature on electroencephalography neglected? And the references in general certainly do not reflect much of the literature in these areas. Finally, in today’s times, when indexes are so easy to do, the book is not sufficiently indexed.

The two editors are respected Brazilian scientists: Alexander Moreira-Almeida is a psychiatrist who has seriously confronted spirituality, and Franklin Santana Santos is a geriatrician with studies about death, dying, and palliative care. Robert Cloninger, a well-known psychiatrist, wrote the Foreword. The contributors, cited above next to their respective contributions, are generally well known. Several of them were recently involved in a symposium at the American Psychiatric Association meeting in May, 2013. This is not surprising as the book’s idea arose out of a symposium.

The book’s theme is well represented by Cloninger’s Foreword (pp. x–xi):

This book on the frontiers of mind-body relationships is a scholarly embodiment of creative and open-minded science. All open-minded people are clearly reminded that strict materialism is a specious and inadequate paradigm—the unhealthy and naked emperor of our scientific era. To restore balance to scientific inquiry, we need only recognize that the consciousness of human beings has a triune nature, one that has developed hierarchically over our long evolutionary history, including procedural learning of habits and skills in our early vertebrate ancestors, semantic learning of symbols and facts in anthropoid apes and early humans, and self-aware learning of narrative language, art, science, and spirituality in modern human beings.
The theme is also well represented by two quotations found in the Editors’ Preface. The first is from Popper and Eccles (p. xiii):

There is a general tendency to overplay the scientific knowledge of the brain, which, regretfully, also is done by many scientists and scientific writers. For example, we are told that the brain “sees” lines, angles … and that therefore we will soon be able to explain how a whole picture is “seen” …. But this statement is misleading. All that is known to happen in the brain is that neurons of the visual cortex are caused to fire trains of impulse in response to some specific visual input.

The second is from Thomas Kuhn (p. xvii):

Scientific revolutions did not triumph because the new paradigm was able to convert all skeptics and leaders of the opposition: The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced. Lifelong resistance, particularly from those whose productive careers have committed them to an older tradition of normal science, is … an index to the nature of scientific research itself. The source of resistance is the assurance that the older paradigm will ultimately solve all its problems.

These exciting ideas put this book in the perspective it deserves. It is a challenge for the mediocre, who may see no further than their nose, and it is an entry into a new era of broader knowledge and consciousness.

References


Vernon M. Neppe

*Pacific Neuropsychiatric Institute*

6300 Ninth Avenue, SE, Suite 353

*Seattle, WA 98115, USA*

*psyche@pni.org*


*Exceptional Experience and Health* is a compendium of research-rich essays that were originally presented as papers in September 2009 at a conference in Liverpool, England (p. 2). Much of the material presented and the tone of the book have a European bent both in the research chosen to be cited as well as treatment programs occurring in the UK and Germany. Indeed the British are way ahead of the US in
having over 15 universities where the study of exceptional experiences and beliefs is readily available to undergraduates, as well as the emergence of several serious academic discussion groups (p. 1). While an informative text, it seems to raise more questions than it answers. The book is divided into two parts with a total of 11 chapters including a helpful introduction by the editor summarizing the content of the book, the various terms used, and a thorough look at the emerging field of clinical parapsychology.

This is not a book for the everyday layperson interested in exceptional experiences on a casual level and looking for personal understanding and useful tidbits to apply in daily life. It is, rather, for the serious researcher interested in a comprehensive overview of the research literature covering each of the various topics included in the book, along with raising questions and an eye to what direction future research should go. Unfortunately, the title may not catch the eye of the population that most needs to read it, that of clinical psychologists who may be conducting research on the impact of prayer or laying on of hands in healing or those working with clients who hear voices or see visions that could fall in the realm of exceptional experiences. I must say that I had been hoping for a book that would bridge the two worlds more easily, so that the lay person could come away with more personally relevant uses and applications of the material presented as it related to their own experiences.

Nonetheless, even as a more research-heavy book, it will hopefully succeed in lessening the gap between the field of parapsychology and that of mainstream clinical psychology with regard to such areas as physical and mental health and exceptional experiences. The key may be in the marketing of the book to universities and researchers as well as, hopefully, future presentations given by the chapter authors at mainstream psychology conferences.

Part One (pp. 27–110) has four chapters containing essays on “Belief, Mind, and Body” with regard to some of the current literature dealing with normal and exceptional experiences and how the mind, beliefs, and experiences can influence not only one’s personal experiences and health but also the experiences and the health of others. Part Two (pp. 113–255) contains the bulk of the book, seven chapters dealing with essays on “Exceptional Experiences and Mental Health.” The field of clinical parapsychology or the more preferred term by some of the authors—“exceptional human experiences” (p.15)—is defined and discussed. This growing field attempts to tease out the difference and shared variance between exceptional experiences and pathology and mental health. Each chapter in the book contains an abstract overview of the essay’s content and highlights critical components and conclusions.

The book begins with Gruzelier’s “The Mind-Body Connection and Healing” (pp. 27–63), which reviews research in general and his specifically on randomized controlled trials assessing the mind-body connection. Gruzelier reports on research dealing with healing the immune system using self-hypnosis, relaxation, guided imagery, hypnosis, Johrei, and Reiki, and begins to tease out personality components that contribute to hypnotizability and positive schizotypy and creativity. His comprehensively reported studies underscore evidence for and validation of use of healing methods that cannot be readily dismissed by skeptics. Practical applications and suggestions include use of healing practices such as Johrei, Reiki, self-hypnosis, and EEG-neurofeedback in helping vulnerable individuals achieve resilience (p.27).

Use of Reiki is already achieving some mainstream recognition. Hospitals in the US are beginning to include holistic medical approaches such as Reiki, aromatherapy, pet therapy, and even massage to help calm patients during presurgery and speed postsurgery recovery. I personally experienced this recently in August 2012 when I had a total right knee replacement. Prominently listed in the hospital’s surgery information packet was the availability of Reiki, aromatherapy, and pet therapy. I availed myself of both Reiki and aromatherapy before and after surgery and felt a personal benefit from their calming and healing impact. The rehabilitation hospital where I then spent a week in postsurgery physical therapy also was introducing the use of Reiki and massage therapy as an option for patients. While Gruzelier covers studies dealing with Johrei and Reiki on health and mental health conditions, surprisingly he does not cover or mention “laying on of hands” or Therapeutic Touch, often used in the US by trained nurses. Nor does he include a provocatively positive outcome study by Bengston and Krinsley (2000) in which mice with mammary adenocarcinoma were healed after a month through “laying on of hands” conducted 1 hour a day for 1 month, by one author, with control mice dying within the expected 14- and 27-day timeframes. This study was replicated three times with the same positive results including using skeptical volunteers.
David Luke’s chapter on “Altered States of Consciousness, Mental Imagery and Healing” (pp. 64–80) explores the anthropological and mythical connections to healing techniques by indigenous peoples and shamans. He looks along the lines of hypnosis, ingestion of psychoactive substances, chanting, dancing, drumming, and breathing, which increase mental imagery. Luke concludes with connections to modern-day psychedelic-assisted psychotherapy in dealing with death-related anxiety and substance abuse (p. 78). Here too it was surprising not to see any mention of Jane Henry’s (2005) edited book Parapsychology Research of Exceptional Experiences with a chapter on healing by Benor and another on shamanism by Krippner.

Williams, Dutton, and Burgess tackle in “Excerpts of Intercoporeality: A Phenomenological Exploration of Energy Healing” (pp. 81–93) how the energetic healing process can involve “effective strategies for changing a person’s lived experience of their illness” (p. 81). The healing process becomes a cocreated experience, whereby the healer and healee develop a joint awareness about the body-mind state of the healee, and offers an alternative stance the healee can utilize. The authors provide theoretical explanations, anecdotal examples, and a case study exploring the use of Chinese energy healing and links to thought field therapy (TFT) and emotional freedom technique (EFT) in Western psychotherapy (pp. 91–93).

This chapter too would have benefited from a more expansive review that included several rich articles dealing with distant healing through use of prayer and an examination of how disclosure of EHEs can result in positive impacts on physical, psychological, and spiritual well-being. In particular, Dossey’s book Healing Words: The Power of Prayer and the Practice of Medicine (1995); Braud’s (1994) article “Empirical Explorations of Prayer, Distant Healing, and Remote Mental Influence”; and Targ and Katra’s (1999) book Miracles of Mind: Exploring Nonlocal Consciousness and Spiritual Healing address significant mind-body experiments and positive distant influences through use of Therapeutic Touch and prayer. Palmer and Braud (2002) offer qualitative and correlational results that show positive and significant relationships between disclosure of EHEs and meaning and purpose in life, well-being, and reduced stress-related symptoms, as well as in spirituality and transformative change. Schmidt’s (2012) meta-analytic research article in the Journal of Alternative and Complementary Medicine showed the positive effect of benevolent intentions in distance healing from 11 eligible studies conducted on three continents. These findings echo those of Astin, Harkness, and Ernst’s (2000) meta-analytic review of randomized trials of distant healing reported in the Annals of Internal Medicine, which found 57% of trials showed a positive treatment effect of distant healing.

Part Two deals with “Exceptional Experiences and Mental Health” (although the title page, 111, mistakenly says “Exceptional Experiences and Health”). The chapter authors do an admirable job of exploring EHEs and their clinical relationship to mental disorder, positive schizotypy, creativity, personal transformation or destruction, and treatment and support options being utilized in Germany and the UK. Here there are many more opportunities for a clinician to walk away with concrete examples, suggestions, and direction in working with clients who share EHEs or ExEs (Exceptional Experiences) that may be “incompatible with their personal and environmental explanation of reality as far as the quality, process, and origin of the experiences (p. 224).”

While Part Two offers much research to back up the various premises and musings, it offers much more qualitative information and examples that make for easier reading over the denser quantitative research presented in Part One. In addition, the material appears to be much more focused toward the clinician than the researcher. However, here too, it was interesting to note that Mack’s (2005) chapter on “Approaching Extraordinary Experiences in the Mental Health Field” in Mijares and Khalsa’s book The Psychospiritual Clinicians Handbook: Alternative Methods for Understanding and Treating Mental Disorders was not referenced, nor was Benor’s (2007) Reaching Higher and Deeper: Workbook for Healing Research, Volume 3—Personal Spirituality: Science, Spirit and the Eternal Soul. A cross-pollination of these authors’ ideas on exceptional experiences and mental health together with the chapter authors’ ideas and explanations of EHEs and mental health in Part Two would have been fruitful in expanding the work of clinical parapsychologists and clinical psychologists treating clients with EHEs.

As a child psychologist and clinical parapsychologist who regularly receives emails from parents about their child’s exceptional experiences, notably having contact with and/or seeing/hearing spirits, I found Belz’s chapter on “Clinical Psychology for People with Exceptional Experiences in Practice” (pp. 223–241) most cogent and useful. She helps to put into perspective and normalize the extent of ExEs (which she prefers to the longer term “exceptional experiences”) in the general population. She also helps
to differentiate between healthy and less healthy ExEs, highlights the overlaps between clinical disorders and applying clinical knowledge to distressing ExEs, and raises some provocative questions for future empirical studies that could help to stimulate and benefit communication between clinical psychology and parapsychology (p. 241). Unfortunately for my work, for the most part, both this chapter, as well as the entire book, focuses only on adults and does not mention or include any discussion regarding children’s EHEs. In part, this is probably due to the significant lack of empirical research studies with children and mostly anecdotal writings on case examples.

Overall, Exceptional Experience and Health offers many gems to the reader. Editor Simmonds-Moore is to be commended for bringing together many expert authors sharing diverse viewpoints, as well as for attempting to rectify the gaps noted above in some of the chapters. In her introduction, she adds another level to the book’s discussions in focusing on her thinking behind the book, along with the relevance of some more recent studies, books, and research, especially in the US (pp. 7–24).

This volume offers much for the professional researcher and even the clinician interested in research and treatment issues. It helps to give the reader a much better understanding of individuals experiencing EHEs as seen through the lens of health and mental health studies and how best they might come together. One can only hope that this book will help in further promoting alternative and nonwestern approaches to medicine, that mental health clinicians will be moved to view the healthy impact of EHEs rather than maintain a pathological viewpoint, and that EHEs will no longer be relegated to just the study and discussions by parapsychologists.

Simmonds-Moore has succeeded in bringing to the reader disparate disciplines that capture how the mind and exceptional experiences can be involved in better physical and mental health. While this volume often raises more questions than it answers, it is well worth including in one’s professional library.

References


Athena A. Drewes

4 Capital Drive
Washingtonville, NY 10992, USA
adrewes@hvc.rr.com
Scholarly research collections of the “paranormal” and the field of parapsychology itself have been focused on phenomenological aspects, theoretical causes, or metaphysical implications. In contrast, *Perspectives of Clinical Parapsychology* is a major attempt to consolidate research in the area of clinical approaches regarding exceptional human experiences. Each of the articles, written by mental health professionals and researchers from Europe and South America, provides an exceedingly valuable perspective and contribution to the knowledge-base. However, the editors and several of the authors rightfully voice concerns about the field as a whole while making appeals for more research. My concerns echo theirs, and I provide a few of my own.

As an important and unique compendium in this field, *Perspectives* is an excellent starting point for interested clinicians. The paucity of clinical research on the needs of experiencers—together with the diversity in approaches, philosophies, and backgrounds—contributes to inherent weaknesses in the relatively new field as a whole. There are divergent views in direction, (i.e. maintaining clinical parapsychology as a “border” field vs. mainstreaming), an overgeneralization of the clinical needs within the “extraordinary experience” population, and the exclusion of spiritual experiences or spiritual factors.

The term “extraordinary or exceptional experience” appears in several articles as an umbrella term for all types of experiences, yet the clinical needs and approaches may be quite distinct, complex, and different. Generalizations are made for the whole population, but the authors have been working or appear to be working with various subsets of experiencers. Distinctions are not always clear. Just as the medical and mental health fields have very diverse subpopulations within greater populations, so are there very distinct subpopulations within the greater population of those of who have had “exceptional experiences.” Different types of experiences most likely will result in different needs and responses. For example, persons who have had a disturbing poltergeist experience will most certainly require a different approach than someone who has had an ecstatic encounter with the divine. A spontaneous, love-filled, near-death experience that is cherished by the experiencer may subsequently result in disorientation, depression, or other post-experience stressors, whereas someone who is dealing with poltergeist activity may want relief from the distressing experiences themselves.

Jon Klimo, in a separate publication entitled “Clinical Parapsychology and the Nature of Reality,” stated that professionals would be optimally trained in “making it [the experience] stop;” or, conversely, “help in finding ways to develop it further for positive use for oneself and others” (Klimo, 1998). The second option is not represented in *Perspectives*.

Additionally, a person who is dealing with an ongoing spiritual emergency, mimicking psychosis, requires yet another set of skills and interventions. “Grandiosity” seen during a spiritual emergency may be a “symptom.” However, for near-death experiencers, “grandiosity” might simply be a normal, excited response to a hyper-real experience with the divine. Kundalini awakenings may require another set of skills. Children and people of other cultures with a range of views of the “paranormal” also have vastly different needs, which adds to the complexity of the field. Without large-scale, diverse, open-ended qualitative and quantitative needs assessments of the many subpopulations, I would argue that very few generalizations or assumptions should be made and that most approaches will eventually need to be subjected to standards not yet developed, confirmed, or tested. For me, this is a matter of ethics.

With the enormous numbers of people reporting these experiences, my view is that this field needs the help of mainstream resources—that is, funding sources, foundations for research, and the support of major institutions. For this, we will need to set very high standards of research and get clearer about the scale and scope of the issues, as well as the populations.

Another weakness in this field is the stripping of spiritual aspects in parapsychological research. From an experiencer’s perspective, spiritual implications may often be at the core of the stressors and the very reason some experiencers seek professional help. Finding purpose in life, dealing with “homesickness” or struggling with new values learned in the experience are just a few of the major challenges faced by those who have had near-death or similar experiences. These issues and how to deal with them are all but absent
from *Perspectives*. Ironically, spirituality itself, while still seemingly invisible to the parapsychology world, is not avoided by the mainstream. NIH has funded hundreds of research projects related to spirituality and health. Resultant positive correlations have led to increases in cultural competency requirements for mental health professionals. The inclusion of spiritual experiences as a nonpathological category in the DSM-IV is another example of increasing mainstream recognition.

Niko Kohl’s chapter entitled “Are Spiritual and Transpersonal Aspects Important for Clinical Parapsychology?” (pp. 135–148) brings the issue of spiritual experiences to the forefront and provides a welcome and much needed historical perspective.

The chapter “A Counseling Approach to Extraordinary Experiences” by Zahradnik and von Lucadou (pp. 118–134) seems to confuse terminology and disciplines when they say people can approach their parapsychological counseling center in Freiburg with “so-called” spiritual experiences, yet the approaches seem to be geared to individuals who have had “occult” or poltergeist-like experiences. Again, no distinction is made. Some comments seem dismissive of the experience itself, invalidating the experiencer, overgeneralizing, and based on opinion rather than objective data. This is hardly an effective counseling approach in my view. Here are a few examples from their conclusions:

Being the victim of unusual activity again also serves as a license for receiving “special help”…. Since most of the time the experiencers still are convinced that they belong to the reasonable, rational part of the population, they often refuse to accept such more outlandish ideas. … Sometimes the experiencers even describe themselves as in some way extraordinary. This measure may be characterized as “occult reorganization.” (pp. 126–127)

Some disturbing experiences such as poltergeist activity may need to be “dedramatized and demythologized [as the] utmost goal of the counseling” (pp. 155–156), as suggested by Bauer, Belz, et al., in their chapter “Counseling at the IGPP (Institut für Grenzgebiete der Psychologie und Psychohygiene),” but other experiences and their responses to them are inherently and understandably dramatic. Some experiences, such as those with spiritual content, distressing or not, may need to be accepted just as they are—as potential catalysts for positive growth. They may also require validation, interpretation, and integration. “Normalization” as a counseling strategy in the sense of offering “natural, physical or physiological and neuro-biological explanations for the clients’ experiences” (p. 156) may be useful for some, but these explanations can also be viewed as invalidation of the experience and therefore harmful. With the absence of scientific evidence accomplished by abundant theoretical and bogus explanations for these experiences, great care should be exercised when attempting to offer explanations. “Normalization” may happen when a client’s experiences are not treated as exceptional, anomalous, bizarre, crazy, or paranormal, but normal. Prevalence studies indicate that as many as 64.9% of the population have had “ESP” experiences and as many as 31.5% of the population have had “numinous” experiences (Levin, 1993).

Another generalization—suggestive of a spiritual experience—in “Theoretical Reflections on Counseling and Therapy for Individuals Reporting ExE” by Belz and Fach, counselors at IGPP, is the following statement. “[Experiencers who have had ‘intense’ experiences] try to avoid difficult negative emotions using evasion towards positive affects as a coping strategy” (p. 183). For a person who has had a deep, spiritual experience, “evasion” of negative emotions might not be a coping strategy, denial, or repression, but part of a deliberately chosen and newly acquired value system. The experiencer may have consciously become part of a growing “culture” that sees negativity itself in an entirely different light. Differences from the “norm” are not necessarily psychological problems. Despite these few concerns, the IGPP serves as a pioneering model of clinical approaches. Their long years of experience in this area have provided significant findings. For example, IGPP saw that understanding the whole person in relationship to the experience is a critical part of an effective therapy (p. 157). In addition, IGPP has provided practice-based evidence that interpretation of metaphorical content in the experience can be an effective therapeutic tool (p. 167).

The shift of focus from the experience itself to the needs of the experiencer requires a leap of faith some scientists are understandably not ready to make. Kramer in “Experiences with Psi Counseling in Holland” points out the issue of traditional research—isolating psi phenomena in order to measure and
control it while the emotional effects of the experiences on the experiencer remain an entirely different matter (pp.7-19). While it is “safer” to focus on the more measurable aspects of the experience itself, in so doing, the deeper meaning of the experience and its impact and potentially positive outcomes for the experiencer can get lost. This shift requires openness and empathy. It requires that the experience be accepted as a profound reality with real consequences for the individual who experienced it—or is experiencing it. If there is doubt about the reality of the experience, the needs of the individual dealing with the consequences of one will be ignored.

If the specific needs and consequences are not well understood, therapeutic approaches may not be as effective or may even be damaging. Holotropic breathwork, for example, may be just what is needed for a person having an ongoing spiritual experience or emergency mimicking psychosis. It is, however, very likely that attempts to re-create a spiritual experience are contraindicated for a person who has had a near-death experience and could already be dysfunctional in his or her obsession to return “home.” This, again, points to the need for more research on the distinctions among different experiencer populations and their specific issues and needs.

There is a huge difference between “parapsychology” and “clinical parapsychology.” Parapsychology is focused on the experience, while clinical parapsychology should be focused on the mental health needs of the person. There may never be proof of a spiritual realm or a reality beyond our own. Even if we can’t measure, imagine, or believe in the phenomenon itself, we can still accept it as reality for someone who is suffering from the direct or indirect consequences of one. As with PTSD, we see the aftereffects of a real experience that we cannot imagine. The phenomenology of the experience itself may play only a part in developing therapeutic interventions, such as when distinctions need to be made between hallucinations and exceptional experiences, or when metaphorical or spiritual content needs to be interpreted as a therapeutic strategy or approach.

Without comprehensive research on the needs and therapeutic approaches of experiencers, clinical interventions are best seen as trial experiments. At such an early stage of research, false assumptions can easily penetrate into treatment systems and are not so easily reversed. Perspectives is aptly named. It should be seen as a collection of points of view. It is not a textbook for clinicians. The significance is in the results of some smaller studies that can stimulate more comprehensive research with various subpopulations. Despite the inherent problems in the field, this book, its authors, and its editors shed much-needed light and awareness. I agree with several authors who urge the establishment of clinical approaches for experiencers as a distinct field of study. The need for distinctions in terminology and demarcation of this field as a distinct discipline is made by Gerding in his article “Philosophical Counseling as Part of Clinical Parapsychology” (pp. 103–117). While not a textbook, Perspectives should be required reading for all who care to venture forth into the complex field of counseling individuals who have issues related to their “exceptional” experiences.

References


Yolaine M. Stout

American Center for the Integration of Spiritually Transformative Experiences (ACISTE)
P.O. Box 1472
Alpine, CA 91903, USA
yolaines@aciste.org
To the Editor:

The review of Neppe and Close’s *Reality Begins With Consciousness: A Paradigm Shift That Works* (RBC) (JP, 76, 395–401) by philosopher Michael Potts has significant differences from the perspective we’ve endeavored to portray. Dr. Potts reviewed the first edition, and over the next four revisions, we have clarified this book greatly. The book is esoteric, and to some is a full undergraduate multidisciplinary course and beyond, and we want to ensure that the statements described are placed in the correct perspective. We quote Dr. Potts (italics) at the start of each of these 22 points and then emphasize our key differences.

The sheer complexity, difficulty, and detail in the encyclopedic first edition (RBC1) makes adequate review formidable for anyone. By the fourth edition, April 2013, RBC has 511 pages, with 496 references, 50 chapters, more than 100 key terms, and enough axioms, theorems, lemmas, and hypotheses to test for centuries, and a special 43-page supplement, *Glimpses and Glossary*, aimed at greater comprehensibility. We’ve carefully revised RBC hundreds of times over 4+ years, refining the ideas into workable science and mathematics and making this, our “life’s work,” more readable, comprehensible, and meaningful. Yet the fundamental axiom and tenets on which it is based remain unchanged, and we have met every challenge put to us, including the Cabibbo challenge below.

The book has received favorable comments from many scientists and scholars in different disciplines, such as (USA) neuroscientist Alan Bachers, psychologist Stan Krippner, consciousness researcher Dean Radin, physicist David Stewart, philosopher Helmut Wautischer; UK biologist John Poynton, Australian parapsychologist Lance Storm, and Canadian mathematician and physician Frank Luger. The Israeli expert on dimensional biopsychophysics Adrian Klein, who has studied RBC intensively over 3 years and critiqued each and every chapter, describes RBC as “... a work that will change mankind’s future.... The 21st Century’s revolutionary paradigm shift ... masterpiece ... monumental ... seismic shift. ... The beginning of the ultimate disclosure about the nature of an all-encompassing reality” (RBC1, p. vi; RBC4, p. vii).

In RBC, we discuss the unique triadic dimensional-distinction vortical paradigm (TDVP). This model generates possibly 600 scientific hypotheses through the key, critically important (a) concepts of triadic space, time and broader “consciousness” (STC) always being tethered together; (b) mathematical and logical “distinctions” of “extent” (measured by “dimensions”) of “content” (container of STC) and of “impact” (influences on content and extent); (c) utilization of a higher level structural content of curved movements—vortices—interfacing across, between, and within dimensions by the process of “indivension”; (d) “paradigms” applied across all the broader sciences—physical, biological, consciousness, and psychological (“metaparadigm”/theory of everything); (e) obedience to the scientific and mathematical laws of nature; (f) unification of the infinite and finite (g) resulting in a philosophical model of unified monism. Yet, Potts does not meaningfully discuss these, largely ignoring (h) the broader sciences; (i) the extensive mathematics; and (j) the validating LFAF (“lower dimensional feasibility, absent falsification”) philosophy of science model, which allows for assessing how higher dimensional domains can be explained in 3S-1t (three spatial dimensions; one moment in time).

Inter alia, Potts mischaracterizes TDVP as “metaphysical,” “neutral monism,” “pantheism,” and “illusion.” Also, he misinterprets the TDVP perspectives of “guiding reality,” “vortices,” “tethering,” “infinity,” “survival,” “transcendence,” and “consciousness.”

We amplify a few responses:

Potts: “Speculative metaphysics.” TDVP is primarily a scientific and mathematical model. It is based predominantly on our combined 60+ years’ study of solid scientific empiricism and mathematical proofs. Applying LFAF, we extend scientific thinking, sometimes suggesting testable hypotheses. For creative ideas, we use “speculative,” “possible,” “guesstimate,” etc. (e.g., we speculate upon the infinite, but not

---

1 E.g., RBC1=First Edition; RBC3+ =Third Edition on. See www.brainvoyage.com
metaphysically—"abstract theory or talk with no basis in reality"). We never even mention "metaphysics": Remarkably, we preliminarily report (RBC4) a profound scientific breakthrough: We can justify the mathematical and scientific feasibility of our finite 9-dimensional rotational (vortical spin) TDVP model (Close & Neppe, 2013). TDVP is only secondarily philosophical, and this is a direct consequence of the science and mathematics portrayed.

Though some of the theories of everything (TOEs) Potts cites are metaphysical, TDVP is not: We compare 24 different TOEs applying detailed, appropriate metrics (general, specific, special): TDVP scores perfectly (39/39), our previous Neppe and Close models >20/39, all others score <20/39!

Potts: "[A]pparently they hold … a form of ‘neutral monism’… being a version of neutral monism, is ontologically parsimonious." Although in the philosophical model of TDVP, unified monism results from scientific empiricism and mathematical modeling, we discuss philosophy extensively: Potts’s comment is incorrect. In neutral monism, both the mental and the physical are reduced to a third neutral substance. Both matter and mind are properties of this unknown substance—two sides of the same coin. Similarly, Potts: "vague pantheism of the universal consciousness view is not clearly defined or defended." Unified monism doesn’t imply neutral monism, pantheism (one God substance), or panentheism (God in all). But TDVP’s “distinctions of impact” allow for many interventions, including guidance. Unified monism developed philosophically from TDVP, with separate but always tethered space/time/consciousness substrates and the finite contained in infinity.

Potts: "Yet the best they can do to describe tethering is through the analogy of ‘bubbles’ in the universal space/time/consciousness." This is a significant misrepresentation: “Bubbles” are not even mentioned: Metaphoric multidimensional balloons are mentioned, but they, too, may mislead. Tethering necessarily involves a relative nonlocality of time, space and consciousness always being together. Previously Minkowski (1908) represented time and space as inseparable; the TDVP axiom argues that time, space and consciousness are inseparable (RBC1): Tethering is far more than multidimensional linkage, implying travel: There’s no travel; tethering is just there. We conceptualize this and other difficult concepts in our supplement to RBC, Glimpses and Glossary. Therefore, in TDVP, space, time, and consciousness substrates always remain profoundly tethered to varying degrees across, between, and within all dimensions (so there is a measure of “extent”).

Potts: “Vortex’ is defined in its usual spatial sense." This is incorrect. In TDVP, a vortex is not just spatial. “Vortex” is a fundamental multidimensional concept involving STC. It is a dynamic moving curvilinear manifold multidimensional distinction of any open or closed form, including spherical, ovoid, helical, or spiral forms (adjective: vortical).

Potts: “… but it is unclear precisely how vortices function in nature and in conscious experience." TDVP’s important model of ‘vortical indivension” creates a multidimensional communication model, a metalevel above field theories. We discuss the concept of vortical indivension in several chapters. Indivension is a process involving fluctuating dimensions of STC—mainly C-substrate domains of “zillions” of individual-units. These portray unique or common transdimensional (often transfinite) relative experiential realities. Indivension occurs through the interaction of vortical distinctions (new term derivation: individual-units; dimensions).

Potts: “The authors’ precluding supernaturalism or a creator God reveals a metaphysical naïveté … natural laws are contingent, requiring a necessary being: …” Egregiously incorrect. We do not exclude God (RBC1 mentioned God 56 times), “infinity of the infinities” (8 mentions), “primary receptor” (40), and guided variants (42)! However, we avoid attributing man-made qualities to these terms: Natural law is fundamental to TDVP. This includes the infinite and guided reality. In our limited 3S-1t reality, we interpret “supernatural” or “miracles,” but at higher dimensional and infinite levels these are not outside nature: Even the incomprehensible, unknown infinite obeys natural laws.

Potts: “Consciousness” is notoriously difficult to define [but Thomas Nagel says that] fundamentally an organism has conscious mental states if and only if there is something that it is to be that organism….”TDVP carefully describes the C-substrate and delineates different kinds of consciousness. Nagel’s is not the ultimate definition for consciousness. We conceptualize different subgroups of consciousness. We refer to C-substrate to properly portray the complexity and unification of consciousness. We devote several clarifying chapters to this topic, particularly from the second edition on, amplifying with each subsequent version. To us, “consciousness” is a broad, general term describing both infinite conscious meaningful information and
Correspondence

finite awareness and responsiveness. These range from the discrete finite physical meaning at inanimate levels and extend to transfinite continuous interactions, modulated in sentient beings in the brain. Quantal (qualit), psychological, neurological, transfinite, and metaconscious levels are pertinent to TDVP, plus philosophical aspects.

Potts: “They are correct that psi operates nonlocally.” Almost but not quite. In TDVP, psi involves relative non-locality not absolute nonlocality. We conceptualize “relative nonlocality” (RBC3+). In consciousness research, “nonlocal”—a distant connection of information, apprehension, or perturbation—is always relative to the observer’s perspective. There are five different levels of relative nonlocality (restricted, dimensional, transfinite, continuous, mystical). Additionally, the nonlocal must express itself relative to the observer, but ultimately it is registered locally (e.g., in consciousness or the brain), hence another reason for the suffix “relative.”

Potts: “But ‘ordropy’ refers only to a tendency toward order—what is ultimately behind the order in living things?” Ordropy is a necessary component of the infinite elements of reality. It reflects spatial, temporal, or other meaningful multidimensional order and patterns, in finite and infinite subrealities, including, but not limited to, negative entropy: Ordropy links closely with our special life model (particularly in RBC2+).

Potts: “But why assume that consciousness, even if defined in terms of some kind of extended information field, operates through some kind of particle?” Potts takes a largely unintelligible quotation out of context, not indicating what came before to make it coherent. He starts in the mid-sentence, ignoring the previous header. This sentence was revised; then in RBC4, it’s completely eliminated. Instead we wrote a new comprehensible Chapter 1 (thank you, Dr. Potts, we listened!), requesting numerous intelligent laypersons to read it, and made it available (at http://www.brainvoyage.com/RBC/excerpts_ch01.php). We simply mention, en passant, a speculative “conscit” model to balance space and time. TDVP does not assume that “consciousness” is necessarily a “particle”: It is a remote theoretical speculation. More pertinent is “DICTU,” the prioritizing description now in RBC4: dimensions, infinity, consciousness, tethering of STC, unification of the infinite and the finite.

With respect to terminology and neologisms, for example, Potts refers to “meta”[meaning] “beyond.” We use “meta” differently, as a prefix implying a broader, higher level of order (e.g., metaparadigm), also in infinite unextended concepts such as metatime, metaspace, metaconsciousness, metainformation and metalife, following on, inter alia, Gödel’s metamathematics and Devereux’s metafamiliarity. Neologisms, all carefully conceptually defined, are necessary linguistically in new multidisciplinary sciences, and particularly in any groundbreaking work.

Potts refers critically to “Individual-units.” Systems theory, which is incorporated within TDVP, is an important, specialized area (e.g., Laszlo). Our applications of compound lengthy words profoundly illustrate such concept unification. We define an individual-unit as a distinct “conscious” finite biological unit across dimensions and also the infinite. Multiple levels manifest together, linked to group, though especially in individuals, and in familial, ethnic, cultural, social, and species contexts.

Potts: “Appealing to a mathematical method, such as Neppe’s co-author Close’s ‘Calculus of Distinctions,’ will not determine which of two or more theories is correct.” This represents an incorrect view of the mathematical representation of reality. Mathematics is not separate from reality. The calculus of distinctions (CoD) is a fundamental mathematical technique which applies mathematicologic. It is no more just a “mathematical method” than is Newton’s calculus. CoD models the processes of the conscious drawing of distinctions, the basis of all perception, conception, and understanding. CoD reflects the logical processes that govern the universe. Its logic therefore applies to any and all conceptual distinctions, including theories. The more a theory conforms to logic, the more likely it is to be correct. The logical structure of a theory has a great deal to “do with” how likely it is to be correct. Using CoD and dimensional extrapolation (RBC1+), we are able to derive from theory results consistent with experimental findings unexplained by the standard physical model. This brings TDVP out of the realm of philosophical speculation into the realm of practical science.

Potts: “In the first place, Close’s calculus should be vouched for in peer-reviewed journals in mathematics and/or logic. If this occurs, it remains the case that mathematical consistency is the mark of many empirically equivalent theories.” We agree. Consequently, every part of every edition of RBC has been peer reviewed, frequently not
just by two reviewers but at times by tens. Every revision has also been peer-reviewed. We have continued
to consider each and every comment received from readers, including Potts’s. Dr. Close’s work has been
reviewed by other competent mathematicians and peer reviewed. Components are published in appropriate
journals or being submitted.

Potts: “The authors’ notion of ‘infinity’ is vague … infinite mind … Where does Neppe and Close’s concept of
infinity fit? It is difficult to tell.” There are several chapters on infinity and models for it, and RBC3 and RBC4
amplify infinity even more. We discuss the transfinite, the finite discrete and the infinite continuum. We
also refer to Gödel and Cantor mathematics. We discuss in great detail the transfinite as opposed to the
infinite and the concepts of continuum versus discrete. It’s a very complex area, even for mathematicians.
The infinite is conceptualized necessarily in terms of a continuum and continuity the finite, in terms of the
discrete. These are key features that are clear and not vague. We don’t use “infinite mind” or “mind.”

Potts: “[I]ndividuality is an illusion? And if “a finite subreality” is pervaded … “how does the infinite
not totally obliterate the finite? If the infinite reality transcends the finite reality, then there is reintroduced a dualism
which Neppe and Close wish to avoid.” Important question: TDVP recognizes reality and does not regard any
component as an illusion. Subjectivity is important, so is objectivity. Individuality is not an illusion. We
talk about “transcendence” but not “infinity transcending”: The many chapters on the topic clarify one
reality: the infinite “containing” the finite. Every finite action is part of the whole infinite reality. There is
no dualism, as there is no separation. TDVP defines “infinite” as limitless, unbounded, continuous, without
end subreality in space, time, and consciousness (C-)substrates; although largely unknown, it obeys the
laws of nature and interfaces with the discrete finite.

Potts: “… put it outside the scope of physics and into the scope of metaphysics.” Contrarily, this is possibly the
most multidisciplinary book ever written! TDVP is far broader than just the specific discipline of physics.
TDVP is likely the single most multidisciplinary scientific and mathematical model ever proposed.

Potts: “What kind of existence do those who survive death have in metaspace? How does time flow for them in
metatime? How is the individuality of their consciousness preserved in their world of metaconsciousness? They would
require intentionality to be capable of seeking communication with the living. Yet if they are part of metaconsciousness,
their individuation as finite beings is difficult to justify.” Good question. This is discussed in detail in RBC2
and amplified in RBC4. TDVP draws some speculative conclusions linking the finite and the infinite. This
includes survival hypotheses.

Potts: “… the more promising of which is James Carpenter’s ‘first sight’ theory of psi.” By RBC3, we had
incorporated the First Sight model into the “distinctions of impact.” This is the major mechanism of
deliberate or automatic psi influence, but other impacts include electrochemical neurological elements,
psychological deliberation with regular conscious or unconscious communications, higher guided
consciousness, bidirectional transfinite and infinite reality elements plus quantal impacts.

Potts: “… by two physicists.” A minor clarification: TDVP was written by two scientists in different
disciplines who complimented each other professionally. Pairing what was pertinent from our backgrounds
facilitated communication of the multidisciplinary needs of the TDVP model. This is relevant as to how
seriously one can take our ideas.

Vernon Neppe, MD, PhD, FRSSAf, DFAPA, BN&NP, is a uniquely qualified and peer-recognized
medical scientist in several disciplines (e.g., the only USA physician ever listed in the peer-selected
America’s Top Doctors in any five different subspecialties, in this instance neuropsychiatry, behavioral
neurology, psychopharmacology, psychiatry, and forensics). He is also internationally peer-recognized in
neuroscience, epileptology, phenomenology, consciousness research, dimensional biopsychophysics, and
parapsychology. (a) As a national, high profile forensic expert on consciousness states and a pioneering
phenomenologist in both consciousness research and neuroscience, he conceptualized a key feature of
the TDVP paradigm as a unified “consciousness” manifesting at every level. With Ed Close, he also (b)
introduced the concepts of a tethered unified triad of space, time, and “consciousness”; (c) proposed the
ubiquity of vortical communications across dimensions; and (d) conceptualized the “continuous infinite”
containing a “discrete finite,” (e) resulting in the philosophical model of unified monism. But Neppe is not
formally a physicist or mathematician, though working in the area for the past decade because of his vertical
dimensional models. Edward Close, PhD, PE, a physicist, mathematician, engineer and cosmologist, used
special mathematics used in TDVP, including: (f) Fermat’s last theorem applied to vortices and asymmetry,
but the solution to this theorem had, in addition, puzzled mathematicians for 4+ centuries and Close was also the first to solve that (Close, 1977); and (g) pioneered the modeling of equations of consciousness and mathematics in Transcendental Physics (Close, 2000) to TDVP. With varying assists from Neppe, he developed, inter alia, (h) “dimensional extrapolation,” a major technique extending geometry to “dimensionometry,” and allowing communications across dimensions; (i) a solution to the Cabibbo angle, which had puzzled physicists for 50 years and consequently the critically important demonstration of the 9-dimensional finite spin model of reality; (j) the calculus of distinctions, which is the most fundamental of all mathematical methods, because distinctions are applied before any other technique, and (k) introduced the concepts of Euclidean and non-Euclidean mathematics in space, time, and consciousness. All these applications of both authors are reflected in the carefully considered mathematics and conceptualizations of TDVP, plus their applying the available empirical data.

In summary, understanding the repetitively updated, multidisciplinary Reality Begins with Consciousness E-book is necessarily nontrivial, because unadorned reality-as-it-is is vast and hypercomplex: Any metaparadigm modeled on reality must be equally so. RBC, containing the TOE of TDVP, endeavors to explain unadorned reality.

References


Vernon M. Neppe and Edward R. Close

Pacific Neuropsychiatric Institute
6300 Ninth Avenue NE, Suite 353
Seattle, WA 98115
tdvp@pni.org

To the Editor:

I appreciate the correspondence of Vernon Neppe and Edward Close in response to my review of their book, Reality Begins With Consciousness. They have put a great deal of effort and research into developing their paradigm and into answering my review, and that speaks for their confidence in the value of their work. In my response I begin by noting the points of agreement I have with the authors. First, I stand corrected in my using the term “bubbles” instead of “metaphoric multidimensional balloons.” I still wonder about the meaning of “tethering” described in that way. Neppe and Close state that “Neologisms work with a whole new area of multidisciplinary sciences, particularly with any groundbreaking work, all carefully conceptually defined.” This is certainly correct—neologisms are important in science when science discovers previously unknown phenomena or needs to develop new theoretical terms. Neologisms must be carefully defined, and ideally they should be used to clarify rather than obfuscate.

I also agree with the authors’ claim that mathematics and reality cannot be wholly separated, a position that dates back to Aristotle. I would note, however, that mathematical and logical consistency may be present in a future competing theory with similar explanatory power to Neppe and Close’s. I stand by
my claim that their theories are underdetermined by the evidence, and this includes mathematical, logical, and empirical evidence.

I am more clear on the authors’ definition of “consciousness,” and I appreciate the clarification. I also grant that as a philosopher who is not trained in the special sciences that I cannot evaluate the mathematical elements of the theory—that is a task for others. It is also true that no review can do full justice to a project as large as Neppe and Close’s—but that does not mean that a review is not helpful in explicating some of the conceptual problems with the theory.

Now to my major points of continued disagreement with Neppe and Close: First, I strongly object to their description of “metaphysics” as “abstract theory or thought with no basis in reality.” One would have thought that such positivism died with the failure of the logical positivist program of eliminating metaphysics during the first few decades of the twentieth century. All attempts to eliminate metaphysics from human discourse presuppose it—and it is not thought “with no basis in reality.” Metaphysics has been variously defined in the history of philosophy, but most would agree that metaphysics has to do with the proper interpretation of reality as a whole. Alfred North Whitehead (1929/1978, p. 3) defined “speculative philosophy” (that is, “metaphysics,” M. P.) as “the endeavour to frame a coherent, logical, necessary system of general ideas in terms of which every element of our experience can be interpreted.” Rather than being pure speculation without regard to rational consistency and empirical evidence, speculative philosophy or metaphysics must be tested by reason and experience; it must be both logically coherent and “applicable” and “adequate” to experience, to use Whitehead’s terminology (p. 3). Given that both metaphysics and science must be rationally coherent and empirically adequate, at the level of fundamental theories of reality, science cannot be separated from metaphysics. Karl Popper (1958/1985) has argued for the criticizability of metaphysical positions. One method is to point out consequences of a metaphysical position and show that they are inadequate—Hume’s atomistic theory of sensation, for example, cannot take adequate account of our perception of wholes. Thus, even if, as I have claimed, Neppe and Close’s system has elements of a metaphysical research program, this would not imply it is irrational, a mere guess, or untrue to experience.

Neppe and Close’s view on science seems similar to “scientism.” This view is also characteristic of some parapsychologists who believe that unless a theory is “scientific” in a narrow sense it cannot be informative of reality, a view most philosophers abandoned long ago but which hangs on in some scientific circles. A research program that cannot be adequately tested for hundreds of years and makes such large-scale claims as Neppe and Close cannot fail to have some metaphysical elements—and that is not a criticism of their theory.

I still fail to see how unified monism would be compatible with the theism of the great monotheistic religions, particularly the doctrine of creation from nothing. That may not, according to Neppe and Close, be a weakness of the theory, but when large scale claims are made about reality as a whole, they do have metaphysical implications regarding the existence and nature of a deity.

Neppe and Close’s system that unifies reality and consciousness gives much food for thought. I am grateful for the authors’ work and for the chance to review it and to respond to their letter.

References


Michael Potts

Methodist University
5400 Ramsey Street
Fayetteville, NC 28311-1498, USA
mpotts@methodist.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 (CRd): 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, unpatterened 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 noncommittal 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

Manuscripts must be submitted to the *Journal* electronically (either as an email attachment or on a CD or disk), preferably in Microsoft Word. The email address is: journal@rhine.org. All submissions should be single column. Preparation of the manuscript should generally follow the guidelines presented in the 6th edition of the *Publication Manual of the American Psychological Association*. However, there are exceptions, so authors should consult the *Journal*’s own guidelines, which are available online at http://www.rhine.org/styleguidelines.html. Articles must include an abstract not exceeding 200 words. Footnotes are discouraged; if they are used, they should be numbered and placed at the bottom of the page. Please pay close attention to references and quotations (see aforementioned *Manual* and the *Journal* guidelines for specific treatment). All items in the reference list should have matching text entries. Please double-check quotations for accuracy and cite page numbers in the text. Each reference list item should be a single paragraph using hanging indentation. See a copy of the *Journal* for the proper format.

Each table and figure should be numbered and have a title or caption in the text, not in a table cell or as a portion of the graphic image. No lines should be used in tables. Tables and figures must be no more than 7 inches wide and should use 11-point type in Times New Roman or a similar font. Figures and photos must be submitted electronically, and they must not be in color. Resolution should be a minimum of 300 dpi. In the case of figures, vector art (e.g., Adobe Illustrator, encapsulated postscript) is preferable to bitmaps. If possible, each figure in its original format should be sent as a separate file named Fig. 1, Fig., 2, etc., in addition to being placed in the desired location in the text file. Figures must often be edited separately and placed in a slightly different location in the text file to fit the final page layout. If figures and photos cannot be submitted electronically, they must be camera-ready.

Manuscripts may be refereed anonymously if so requested by the author.

The *Journal of Parapsychology* has the following additional guidelines:

1. State the precise statistical formulations of any hypotheses and list them before the results.

2. Report not only inferential statistics (e.g., t values), but also the descriptive statistics for the data evaluated (e.g., group means and standard deviations). Also, present the actual values of correlation coefficients, not merely whether a correlation is significant.

3. Arrange to have data and statistical analyses independently rechecked before submitting a paper.

Manuscripts accepted for publication are copyedited for grammar and style. Changes of consequence are submitted to the author for approval, and proofs are sent for corrections. A prompt response is greatly appreciated. Extensive revisions at this stage are charged to the author.

Articles and reviews in the *Journal* must be original, which means they must not have been published previously, either in whole or in part, including on 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 elsewhere, including on the Web.