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

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

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

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 Lexscience.

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 © 2014 by the Parapsychology Press
ISSN 0022-3387
INVITED ARTICLE

Star Gate: The U. S. Government’s Psychic Spying Program

Edwin C. May

ARTICLES

The Nature of Precognition

Jon Taylor

Revealing Psi Secrets: Successful Experimenters Seem to Succeed by Using Their Own Psi

Adrian Parker and Brian Millar

Beyond the Coin Toss: Examining Wiseman’s Criticisms of Parapsychology

Johann Baptista and Max Derakhshani

How to Remove the Influence of Expectation Bias in Precognition Experiments: A Recommended Strategy

Jan Dalkvist, Julia Mossbridge, and Joakim Westerlund

Mediumship, Psychical Research, Dissociation, and the Powers of the Subconscious Mind

Carlos S. Alvarado

Precognitive Dreaming: Investigating Anomalous Cognition and Psychological Factors

Caroline Watt

OBITUARY

Robert L. Van de Castle

By Erlendur Haraldsson
BOOK REVIEWS

*Education in Parapsychology: Student and Instructor Perspectives*
By Harvey Irwin
*Chris A. Roe*

*Conversations With Ghosts*
By Alex Tanous and Callum E. Cooper
*Pamela Heath*

*Randi’s Prize: What Sceptics Say About the Paranormal, Why They Are Wrong, and Why It Matters*
By Robert McLuhan
*Graham Watkins*

Editor’s Note: Policy on Author Pseudonyms

Glossary

Instructions for Authors

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

STAR GATE: THE U.S. GOVERNMENT’S PSYCHIC SPYING PROGRAM

BY EDWIN C. MAY

ABSTRACT: From 1972 to 1995 various agencies of the U.S. government funded applications of and research into psi to the tune of over 20 million U.S. dollars. Although this sounds like a substantial amount of money to most of us, with regard to military and/or intelligence funding it is almost round-off error! This activity was not inspired by some academic curiosity that one might find at a university; rather, it was driven by necessity during the Cold War. Puthoff and Targ deserve unbounded respect for shepherding the project, especially in its early days. My view of STAR GATE extends from late 1975 through 1995, and I am the “keeper of the keys” of all the research and some of the spying. This means I have all the project records, including such things as raw data from a large number of experiments, final reports to a host of clients, administrative records as to who funded the project and for how much money, who was involved, and how and why the government’s in-house activity was established at Ft. Meade. This paper is a personal narrative of my first-hand account of much of that work.

Keywords: psychic spying, government ESP, insider view, intelligence operations

It was late in 1975, in the midst of the Cold War, that I joined the ongoing, highly classified program at Stanford Research Institute (now called SRI International) as a consultant to study and use extrasensory perception, a form of which technically is also known as “remote viewing,” to gather information for the U.S. military and intelligence communities. After I had been notified that my secret government clearance had been approved, the director of the program, Harold Puthoff, called me into his office and closed the door behind me. Hal opened an imposing safe in his office. What he showed me blew my mind. Even to this day, after all these years, I still get goose bumps thinking about the then-classified examples he showed me.

Government clearances were not new to me. I was given my first one as a 20-year-old student while working during the summer at the Rand Corporation in Santa Monica, California. I worked there for each summer for the next 5 years in the Earth and Planetary Sciences Department learning about atmospheric physics. Physics, after all, was my declared major at the University of Rochester, and this job offered me an extraordinary opportunity in theoretical research.

The data Hal took from his safe have all been declassified and are now available from the National Archives for $140 (www.archives.gov). Thus, I will not belabor the details here. I simply quote from Puthoff and Russell Targ’s final report to the CIA, dated December 1, 1975:

Date: 1 June 1973, 1700 hours, Menlo Park, California.

Protocol: Coordinates 38°23’45” to 48°N, 79°25’00”W were given (with no further description) by experimenter Dr. H. E. Puthoff to subject SI [later identified as Pat Price] by telephone to initiate the experiment.

On the morning of 4 June 1973, SI’s written response (dated 2 June 1973, 1250 to 1350 hours, Lake Tahoe, California) was received in the mail:

Looked at general area from altitude of about 1500 ft above highest terrain. On my left forward quadrant is a peak in a chain of mountains, elevation approximately 4,996 ft above sea level. Slopes are greyish slate covered with variety of broadleaf trees, vines, shrubbery, and undergrowth. I am facing about 3° to 5° west of north. Looking down the mountain to the right (east) side is a roadway—freeway, country style—
curves then heads ENE to a fairly large city about 30 to 40 miles distant. This area was a battleground in
civil war—low rolling hills, creeks, few lakes or reservoirs. There is a smaller town a little SE about 15 to
20 miles distant with small settlements, village type, very rural, scattered around. Looking across the peak,
2500 to 3000 ft mountains stretch out for a hundred or so miles. Area is essentially wooded. Some of the
westerly slopes are eroded and gully washed—looks like strip mining, coal mainly.

Weather at this time is cloudy, rainy. Temperature at my altitude about 54° F—high cumulo nimbus
clouds to about 25,000 to 30,000 ft. Clear area, but turbulent, between that level and some cirri stratus at
46,000 ft. Air mass in that strip moving WNW to SE.

1318 hours—Perceived that peak area has large underground storage areas. Road comes up back side
of mountains (west slopes), fairly well concealed, looks deliberately so. It’s cut under trees where possible—
would be very hard to detect flying over area. Looks like former missile site—bases for launchers
still there, but area now houses record storage area, microfilm, file cabinets; as you go into underground
area through aluminum rolled up doors, first areas filled with records, etc. Rooms about 100-ft long, 40-ft
wide, 20-ft ceilings, with concrete supporting pilasters, flare-shaped. Temperature cool—fluorescent light-
ed. Personnel, Army 5th Corps Engineers. M/Sgt. Long on desk placard on grey steel desk—file cabinets
security locked—combination locks, steel rods through eye bolts. Beyond these rooms, heading east, are
several bays with computers, communication equipment, large maps, display type, overlays. Personnel,
Army Signal Corps. Elevators.

1330 hours—Looked over general area from original location again—valleys quite hazy, lightning
about 30 miles north along mountain ridge. Temperature drop about 6° F, it’s about 48° F. Looking for
other significances: see warm air mass moving in from SW colliding with cool air mass about 100 miles
ESE from my viewpoint. Air is very turbulent—tornado type; birds in my area seeking heavy cover. There
is a fairly large river that I can see about 15 to 20 miles north and slightly west; runs NE then curves in
wide valley running SW to NE; river then runs SE. Area to east: low rolling hills. Quite a few Civil War
monuments. A marble colonnade type: “In this area was fought the battle of Lynchburg where many brave
men of the Union and Confederate Armys [sic] fell. We dedicate this area to all peace loving people of the
future—Daughters G.A.R.”

On a later date Sl was asked to return to the West Virginia site with the goal of obtaining information on
code words, if possible. In response, Sl supplied the following information:

Top of desk had papers labeled “Flytrap” and “Minerva”. File cabinet on north wall labeled “Operation
Pool” … [third word unreadable].

Folders inside cabinet labeled “Cueball”, “14 Ball”, “Ball”, “8 Ball”, and “Rackup”. Name of site
John C. Calhoun (?).  

Urals Site (S1)

After obtaining a reading on the West Virginia site, Sl volunteered that he had scanned the other side of
the globe for a Communist Bloc equivalent and found one located in the Urals at 65°00’57”N, 59°59’59”E,
described as follows:

Elevation, 6200 ft. Scrubby brush, tundra-type ground hummocks, rocky outcrops, mountains with
fairly steep slopes. Facing north, about 60 miles ground slopes to marshland. Mountain chain runs off to
right about 35° east of north. Facing south, mountains run fairly north and south. Facing west, mountains
drop down to foothills for 60 miles or so; some rivers running roughly north. Facing east, mountains are
rather abrupt, dropping to rolling hills and to flat land. Area site underground, reinforced concrete, doorways
of steel of the roll-up type. Unusually high ratio of women to men, at least at night. I see some helipads,
concrete. Light rail tracks run from pads to another set of rails that parallel the doors into the mountain.

Thirty miles north (5° west of north) of the site is a radar installation with one large (165 ft) dish and
two small fast-track dishes.

The two reports for the West Virginia site and the report for the Urals site were verified by personnel in the sponsor
organization as being substantially correct. The results of the evaluation are contained in a separate report filed with the Contracting Office Technical Representative.

As it turned out, the West Virginia site was a very secret National Security Agency (NSA) listening post, and S1’s data spawned a substantial internal security investigation that showed no wrongdoing on the part of the SRI team or S1. All this happened before the Internet and tools such as Google Earth. In essence, it was data such as these from S1 and others during those early days of the project that cemented the U.S. government’s commitment to the remote viewing programs for the next 20 years.

From the inception of the project under the CIA’s auspices in 1972 through 1979, SRI had three primary responsibilities. First, we were to use ESP to obtain information about potential threats from the Soviet Union, other Eastern Bloc nations, and the People’s Republic of China. Two, we assessed the credibility and accuracy of intelligence regarding ESP research that was slowly filtering out from the Soviet Union. Finally, with minimal support, we conducted basic and applied research. The basic research concerned the fundamental physics, physiology, and psychology of ESP. The applied research searched for ways to make the “end-product” more accurate and reliable.

It is a sad fact that modern military decision makers are extremely hesitant to finance programs based on a putative extrasensory capability. During the Cold War, Senator William Proxmire invented a prize—the Golden Fleece Award—as a way of embarrassing government officials who routinely funded silly projects. The study of ESP also possessed a high “giggle” factor, regardless of the quality of the work. Both the giggle factor and the fleece award had a chilling effect on the funding community for ESP research. When I became the project director at SRI, more than 40% of my time was spent attempting to raise funds so that the program could continue.

There were many successful applications of ESP within the project at SRI and later at Science Applications International Corporation (SAIC). SAIC is similar to SRI in that it is a private, not-for-profit corporation, but it is much larger. Although it has a number of nondefense contracts, it is, or was then, primarily a defense contractor. When the program closed at SRI for lack of funds, a former Air Force client, then retired, was working for SAIC and arranged for me to establish STAR GATE anew there. That lasted for another 4 years with funding of about four and one half million dollars.

One example of an intelligence-like success. Consider the case of trying to find U.S. Army Brigadier General Dozier, who was kidnapped from his home in Verona, Italy on the evening of December 17, 1981. Joe McMoneagle was asked to locate the general by using remote viewing to accurately draw Dozier’s current location. Among McMoneagle’s responses was a drawing of a unique circular park with a cathedral. As it turned out, by scouring maps and photographs for such a combination of structures, the searchers found one in the city of Padua, the place where General Dozier was later rescued. Later Dozier was briefed on February 9, 1982, from 9:30 to 10:15 a.m., at the Command’s Special Compartmented Information Facility (SCIF: Bldg. 4554) on GRILL FLAME an early nickname for the psychic program. He was then asked to review sketches and narratives generated during the GRILL FLAME sessions for any correlation to places or events surrounding his kidnapping. Dozier was so impressed with the data that he suggested that senior government officials, military officers, and leading business and political personalities be instructed on what to “think” if they were kidnapped so that psychic searchers could more easily locate them.

There are three basic approaches that have worked for such search tasks in the field and in the laboratory. The first is to ask psychics to “stick a pin in a map” corresponding to the lost person. In my experience, this approach usually doesn’t work. When it does work, people often give this positive result undeserved attention.

A more effective method is the standard “remote viewing” technique, where the target is the whereabouts of the missing person, in this case General Dozier. However, even this approach has its problems. Excellent remote viewing might not contribute much to finding the lost person. Let me illustrate. Imagine the following scenario. We wish to find a Soviet submarine that is lurking underwater somewhere off the California coast. Fortunately, we have at our disposal a psychic viewer who is nearly perfect with her impressions. The viewer describes the interior of the sub exactly, describes the crew members in detail, provides the name of the captain and his children, and tells what the crew ate for dinner that evening! We now have top-of-the-line accurate psychic data, but it in no way helps us find the sub. Then, using remote viewing to “look outside” the sub yields an amazingly accurate description of—you guessed it—water! This is one example of how intelligence value is often unrelated to the quality of the remote viewing. The quality of remote viewing is excellent; the value of the information nil.

Fortunately, the real world provides a compromise. In the standard out-bound remote viewing protocol, an agent travels to some randomly chosen location and the viewer simply describes the surroundings where that person
Although the bad news should be construed as good. Because we were good at our jobs, we won 85% of the vigor-presented our results and discussed the outcomes with the group in person. There was good and bad news here, with our group at SRI. Their opinions were added to our final reports as appendices.

I was transferred to the Army Presidio in San Francisco, but whose full-time responsibility was to be in his office. Contracting Office Technical Representative. In our case, this was an Army officer with the rank of colonel who had

The member was editor-in-chief. They took notes and eventually provided their comments in writing directly to the SOC members. They were to review them as if they had been submitted to a scientific journal of which the reports were completed and copyedited by SRI International professional editors, they were copied and sent

The quality of their reviews, I can attest that the quality of our output improved substantially.

One aspect of the large Army contract was that we were required to conform to the wishes of three separate Army-constituted panels: a Scientific Oversight Committee (SOC), an Institutional Review Board (a.k.a. Human Use Review Committee), and a Pentagon Policy Review Committee. All committee members were required to hold active security clearances. Let me emphasize that these committees were not “rubber-stamp” bodies. Rather, their members agreed to long-term commitments, and they all took their responsibilities very seriously. As the recipient of their reviews, I can attest that the quality of our output improved substantially.

Probable the most active committee was the SOC. For the first 2 years, it had 12 members who were drawn both from lists of people we supplied and ones provided by the Army. Because in the third year of the contract the 2 million dollar budget was cut in half, I reluctantly had to reduce the SOC membership to only five. The Army had the final decision regarding who would serve on the committee. They all were paid under our contract for their time and travel. A threshold requirement for serving on the SOC was that the member be skeptical of putative parapsychological phenomena, but at the same time open-minded enough to want to take the job seriously. Furthermore, their time commitment was substantial, as the job was to last for the 5-year duration of the contract.

The SOC had three primary tasks: (a) to review and approve the detailed experimental protocol for every experiment to be conducted under the Army contract; (b) to exercise unannounced drop-in privileges to see firsthand what was happening; and (c) to review critically, in writing, the final reports for each of the tasks in the contractual statement of work. There were 38 of the latter in the first year alone.

Because our group was highly professional, the first of the SOC’s three tasks was rather straightforward. From time to time they did some protocol “tweaking,” but, for the most part, the protocols we submitted were approved directly, with little or no substantive change. The SOC’s second task—unannounced drop-in privileges—looked good on paper but was hardly ever exercised. I suppose this was to be expected, given that the committee members were senior professionals with active individual careers.

The main SOC action came with their third responsibility—critical reviews of our final reports. As soon as the reports were completed and copyedited by SRI International professional editors, they were copied and sent to all the SOC members. They were to review them as if they had been submitted to a scientific journal of which the member was editor-in-chief. They took notes and eventually provided their comments in writing directly to the Contracting Office Technical Representative. In our case, this was an Army officer with the rank of colonel who had been transferred to the Army Presidio in San Francisco, but whose full-time responsibility was to be in his office with our group at SRI. Their opinions were added to our final reports as appendices.

In addition to obtaining their written opinions, we hosted each year a 2-day meeting, during which we presented our results and discussed the outcomes with the group in person. There was good and bad news here, although the bad news should be construed as good. Because we were good at our jobs, we won 85% of the vigorous and sometimes loud arguments, but the better news is the 15% we lost. Our scientific “product,” so to speak, sharply improved. That improvement manifested in two ways. First, we learned to approach all positive results from our experiments in a skeptical way; that is, we learned to assume that what we just saw was a mistake and we needed to set about finding it. If we failed to find an error, we could assume something interesting was happening. Secondly, because of the interdisciplinary nature of the SOC, our group was exposed to experimental and theoretical
techniques that were outside the training of our own researchers but could be incorporated with the SOC’s assistance. My interaction with the SOC has been among the highlights of my professional academic research career.

Besides direct ESP, here was another approach to gathering intelligence in which the U.S. government took an interest. Normally, when a new military policy, weapons system, or battle order is being considered, the proposed new system is evaluated critically. Often two teams of evaluators, designated as Red and Blue, are assembled to criticize or support the plan. Our group was awarded a contract to participate as part of a Red team to evaluate a proposal by the Carter Administration, and later by the Reagan Administration, to deploy the new MX missile system.

The proposals were variations on the theme of building many more missile-launching facilities than there are missiles and then continually moving the missiles covertly among the various launching facilities—let’s call it a nuclear shell game. The Congressional Budget Office originally estimated that procuring 200 MX missiles, building 5,800 shelters for them, and operating the system would cost 28.3 billion dollars all the way through to fiscal year 2000. Eventually, a “racetrack” idea gained favor. Each missile would be moved among shelters located on spur roads radiating from a central, circular track. There would be about five such patterns, or clusters, in each of about 40 valleys in the deserts of Nevada and Utah. This racetrack system would allow transporters to shift missiles between shelters within 30 min, in time to escape incoming Soviet missiles after they have been launched. The racetracks would be about 56 km in diameter.

The complexity and financial support that would have been required for this proposed system demonstrates how seriously the Carter Administration considered the concept. In fact, they were planning to move ahead as soon as possible. The U.S. Air Force expected to begin site selection for the MX operation’s base test and training facility by 1980. Work on the first racetrack and shelters would begin by 1983. The first 10 MX missiles and 230 shelters were scheduled to be operational by 1986. The assumption behind this system was that the Soviets would not know where to aim their missiles to cause the most damage to the US’s ability to retaliate, as our missiles would be moved continually among the various shelters.

The question before us was whether we could compromise this racetrack concept using ESP. If so, we would have to assume that the Soviets could also accomplish this. This would imply that the racetrack concept was indefensible, and it would make no sense to build this vast system of racetracks and shelters in the first place.

Our proposal, which was eventually approved, included the following elements in the statement of work, which indicated what we were going to do with the money if the contract was awarded to SRI:

1. As a feasibility study, we would assume 1 actual missile shelter to 9 sham ones, determine the statistics of an MX system compromise as a function of beyond-chance hitting by remote viewers.
2. Conduct a screening program involving about 100 SRI employees and other experienced remote viewers, utilizing a 1-in-10 screening device with 100 trials for each of these employees.
3. Take the five best people from the screening and have each of them contribute 200 more trials.
4. From these data, estimate the potential vulnerability of the racetrack or shell game concept.

In addition, we used a sophisticated statistical technique coupled with a form of ESP called “dowsing” to see if we could compromise the system by using a statistical approach towards compromising the MX missile system.

In our final report to the government, we showed that ESP practitioners were able to locate the hypothetical missile in 12 out of 12 trials, with a total of 452 circle selections (May & Puthoff, 1981). The correct hit rate was over two-and-one-half times what was expected in a 1-in-10 game. Figure 1 shows a letter on U.S. Senate stationery from Senator John W. Warner (R-Virginia) to the Secretary of Defense at the time, the Honorable Caspar W. Weinberger, describing our contribution to the MX missile program.

I doubt that our data alone kept the system from being built, but, on the other hand, our ESP research reports surely contributed in an important way towards that end.

Beginning in 1986, the Air Force became exceptionally interested in learning the degree to which remote viewing could provide useful information on directed-energy weapon systems. To test this idea, they awarded us a contract to examine this question in three trials: one per year, for 3 years. As always, we used a double-blind protocol, meaning that no one who interacted with the remote viewers knew anything about the potential target or even, in this case, the identity of the client. A session would play out as follows. We were usually given the Social Security number of an individual none of us had met. In addition, we were told that on a specific date this person would be
somewhere in the continental US. As project director, I knew that the targets would be directed-energy systems of some kind, but beyond that I too did not know any specifics.

At a specified time, Nevin Lantz, our project’s psychologist and active researcher, would assign a task to the psychic at midnight and again once every 8 hours, including the next day’s midnight. That task was simple: describe the surroundings where the person to whom the Social Security number belonged was standing. So far, nothing particularly new or inventive was involved. The analysis of the result was a breakthrough not only for laboratory studies. After all, if used properly, it could easily have been adapted to the real world of psychic spying.

I will not go through the mathematical intricacies, because conceptually the idea is quite simple. Before any of the sessions with a client began, I worked with the sponsor to define three categories of things they wanted

Figure 1. Letter from Senator John W. Warner commenting on the success of our MX missile program compromise.
to know about the target. First and foremost was the target’s function: why it was being developed. The Air Force had five or six different functions in this category alone in which they were interested. The second category was physical relationships: an instrument, for example, might be underneath a building which is next to a truck. There were around 10 such relationships. Lastly, they specified a rather long list of objects, similar to those one would expect from a traditional remote viewing.

For each of the targets, the Air Force filled out a table for each element in all three categories. The elements in the table were specific to each target to be employed later, and it included ratings of the degree to which each element was germane to that target. After the psychic session, an analyst, who was blind to the target and its list of items, filled out the same table, but this time with regard to the degree to which each item was present in the psychic’s response.

Armed with both tables, one for the intended target and one for the response, the computer could take over. Although mathematically complex—the process is known as “fuzzy set analysis”—three simple ideas emerged from the computation. Accuracy was defined as the percentage of the Air Force’s predefined target elements that was obtained by the psychic. The reliability was defined as the percentage of the elements in the psychic’s response that were correct. Finally, the “figure of merit” was defined as the product of accuracy and reliability.

The way, then, to obtain a high figure of merit was for the psychic to describe as much of the intended target as possible, but in as simple and concise a way as possible, so as not to include many incorrect elements. To get a hint of what a random response could be like in the absence of any psychic ability, we had determined in the laboratory that, using a rough rule of thumb, about a third of any site could be described by about a third of any response. This may seem high, but this rule of thumb arose from considerable analyses of data collected in the laboratory.

How did this work out? I will describe just one of three successful examples: Project Rose, a high-frequency, high-power, microwave device in the New Mexico desert at Sandia National Laboratory. Joe McMoneagle was the psychic on this trial. By the Air Force’s own assessment, the accuracy, reliability, and figure of merit for this case were 80%, 69%, and 55%, respectively. Keep in mind that chance, Joe just being lucky, would predict these numbers to be 33%, 33%, and 11%, respectively. Figure 2 shows that the drawing and pictures were more impressive.

From my 30 years of experience in ESP research, I consider this case to be among the very best. If this example had been an intelligence operation instead of a proof-of-principle session, an independent analyst would have had no trouble whatsoever in identifying the target as a microwave device of some sort. The drawings on the bottom in Figure 2 clearly show easily identifiable elements, such as a waveguide and microwave horn. Joe went on to say that this device was in a wrapped environment and was being used as some kind of test evaluator. In fact, they were shining the microwaves on electronic instruments to test their sensitivity to high-energy microwave radiation. For me, however, the pièce de résistance is the drawing in the upper left side of Figure 2. Perhaps it is not as visually compelling as other examples, but for me, at least, “He nailed it!” Not only did Joe accurately describe exactly what was going on, but also by his drawing he indicated the spread of the electromagnetic radiation, which matched the known beam angle of the device. The actual device is shown at the top of the figure.

My point in all this detail is important. We had developed a system of analysis that had the potential of allowing an operations analyst looking at real psychic spying data to evaluate the results quantitatively. When combined with more traditional methods of intelligence collection, this method allowed the military to assess more accurately whether or not to invest resources in solving the problem.

Did the psychic spying program work? In short, Yes. I realize, of course, that the “official” U.S. government’s response was “No.” However, as I hope I have demonstrated here, the real answer is more complex. The CIA was tasked by Congress in 1995 to conduct a 20-year retrospective review of STAR GATE to determine whether or not the intelligence community should continue funding the program. They hired the American Institutes for Research to conduct the investigation, and they produced two reports: one classified and one not (Mumford, Rose, & Goslin, 1995).

In a release of many of the STAR GATE documents in 2000, the CIA published a report entitled “Summary Report: STAR GATE Operational Tasking and Evaluation in which they conducted a detailed analysis of 40 ESP operations. Quoting from this report (Mumford, Rose, & Goslin, 1995):

From 1986 to the first quarter of FY 1995, the DoD paranormal psychology program received more than 200 tasks from operational military organizations requesting to attain information unavailable from other sources. The operational tasking comprised “targets” identified with as little specificity as possible to avoid “telegraphing the desired response.”
In 1994, the DIA STAR GATE program office created a methodology for obtaining numerical evaluations from the operational tasking organizations of the accuracy and value of the products provided by the STAR GATE program. By May 1, 1995, the three remote viewers assigned to the program office had responded, i.e., provided RV product, to 40 tasks from five operational organizations. Normally, RV product was provided by at least two viewers for each task (p. D-1).

Data from these 40 operational tasks were evaluated by the tasking organization (not by the ESP team members) along two separate dimensions. About 70% of the 100 separate evaluations of these data were deemed to be possibly true or better; however, only 50% were deemed to be of some practical value, however minimal.
Before we jump to conclusions that the spying unit was worthless, there are a number of major problems not mentioned in this particular CIA report that we must consider. First, the evaluations shown above were all gathered “after the fact”; that is, after some form of “ground truth” had been determined. In all the years of research effort under the STAR GATE program, we were never able to identify in advance a reliable indicator of the value of the data in a particular response, in total or in part. Thus, it would be considered a major risk to assign scarce resources to intelligence gathered by ESP without having confirmatory data from other independent sources and methods.

The conclusion was that ESP was not particularly useful, so the CIA eventually decided not to assume responsibility for the STAR GATE program in 1995, although they did suggest that the academic community continue to look into ESP—an odd comment if they had been convinced that ESP did not actually exist. Thus, the government sponsorship of ESP activity came to a close.

It was clearly a mistake to curtail our operations, based on their analysis. By the CIA’s own admission, they only evaluated 40 sessions out of many hundreds, and they only looked at data from 1994 onward. Even though they were requested to do so, they did not interview McMoneagle or any of the individuals who were responsible for his receiving a Legion of Merit award—the highest honor for any intelligence officer—for his excellent contribution to intelligence collection. To place this into context, allow me to quote from a part of the citation for this prestigious award:

…[McMoneagle] used his talents and expertise in the execution of more than 200 missions, addressing over 150 essential elements of information. These EEI contained critical intelligence reported at the highest echelons of our military and government, including such national level agencies as the Joint Chiefs of Staff, DIA, NSA, CIA, DEA, and the Secret Service, producing crucial and vital intelligence unavailable from any other source….

This inconvenient citation was never considered in the CIA decision, nor was it part of the overall investigation ordered by the U.S. Congress to evaluate a 20-year-long program. The issue of whether the unit was pulling its own weight in the intelligence community remains murky.

According to McMoneagle’s assessment, during his time at Ft. Meade from 1978 to when he retired in 1984, approximately 15 to 20% of the cases of psychic espionage were resolved successfully. This sounds terrible. But let us remember that the program, first at SRI and later at Ft. Meade, always seemed to be a court of last resort. Only the “impossible” problems were tasked, that is, those problems that did not yield to traditional methods of intelligence collection. Thus, from that perspective, a 15 to 20% success rate is as close as one can get to a miracle. Many of these successes remain classified, along with the few that been gained since then. Here I may also add the oft-quoted “intelligence failures” that occurred when traditional human and technological intelligence methods were used after the recent spate of terrorist attacks in many parts of the world. The point is that we have to ensure we use all available sources of data collection to decrease the instances of intelligence failure.

Shortly after the government closed the Ft. Meade Unit, Congress required the staff to send all their records to the CIA. The person who helped pack up the material told me that many of our research reports were still in their original packages, unopened! This is one terrible consequence of self-defeatism and of a well that was poisoned against scientific inquiry.

In the closing days of the Ft. Meade Unit, the staff sent to the CIA approximately 35 sealed boxes so that the agency could conduct their Congressionally-directed evaluation of the STAR GATE program. In both the classified and unclassified versions of their report to Congress, they implied that the result of their careful examination of the record showed that further military or intelligence community support was not warranted.

Two years later, after the CIA program evaluation reports had been published, two colleagues, one from the DIA and one from the Pentagon, officially were given access to the room at the CIA in which all the boxes were stored. They found that not a single sealed box had ever been opened! The fix was in!

The implications are obvious: the careful and in-depth review of the material required by the U.S. Congress was not based on an evaluation or even a reading of the data! Because the DIA person had helped pack the boxes and could identify which ones to open, in a matter of minutes they were able to find incontrovertible proof of intelligence collection examples that were not only successful, but constituted a valuable contribution to solving the problem at hand. Former CIA Director Robert Gates commented on the news television program ABC Nightline (1995):
Well, all I can say is that in the 20 years or 25 years where I was in a position perhaps to be aware, I don’t know of a single instance where it is documented where this kind of activity contributed in any significant way to a policy decision or even informing policymakers about important information.

This statement is blatantly false, and Director Gates was aware that I knew he had been briefed on specific examples to the contrary. In fact, my role on this episode of Nightline was simply to act as a foil to Director Gates. Many of my comments contradicting him were edited out in the final aired version. I chose to go on this program, even though my managers at SAIC ordered me not to, and I was issued the threat, by implication, that both CIA and SAIC lawyers were going to watch the show for any transgressions I might commit. My only alternative was to resign my post at SAIC, effective immediately.

Under a contract from the CIA, Mumford et al. (1995) prepared a report for the American Institute for Research on the STAR GATE program. This report was based on the expert evaluation of Professors Jessica Utts (1995a, 1995b), dubbed as representing the pro-psi group, and Ray Hyman (1995), considered a skeptic. The tasking for the reviewers was to answer four general questions: (a) Was there a statistically significant effect? (b) Could the observed effect, if any, be attributed to a paranormal phenomenon? (c) What mechanisms, if any, might plausibly be used to account for any significant effects and what boundary conditions influence these effects? (d) What would the findings obtained in these studies indicate about the characteristics and potential applications of information obtained through the remote viewing process? As Mumford et al. (1995) reported:

One of Dr. Hyman’s first comments about Dr. Utts’ review was that he considered it perhaps the best defense of parapsychological research he had come across. We concur; likewise, we feel that Dr. Hyman’s paper represents one of the clearest expressions of the skeptic position we have seen.

At the outset, it should be noted that the two reviewers agreed far more than they disagreed. One central point of agreement concerns the existence of a statistically significant effect: Both reviewers note that the evidence accrued to date in the experimental laboratory studies of remote viewing indicate that a statistically significant effect has been obtained. Likewise, they agree that the current (e.g., post-NRC review) experimental procedures contain significant improvements in methodology and experimental control (p. 3-80).

Utts (1995a), a professor of statistics at the University of California at Irvine, and formerly at Davis, stated as part of her analysis:

Using the standards applied to any other area of science, it is concluded that psychic functioning has been well established. The results of the studies examined are far beyond what is expected by chance. Arguments that these results could be due to methodological flaws in the experiments are soundly refuted. Effects of similar magnitude to those found in government sponsored research at SRI and SAIC have been replicated at a number of laboratories across the world. Such consistency cannot be readily explained by claims of flaws or fraud (p. 3-2).

The second expert reviewer, Hyman (1995), a professor of psychology at the University of Oregon, while agreeing on the statistical evidence, felt that competing explanations of the phenomena had not been eliminated, a viewpoint that is disputed by Utts (1995b). However, considering these are early days in the investigation of psi phenomena, and there is a growing body of research from other groups and laboratories, in the spirit of science and inquiry we need to continue with this investigation and attempt to address the most vital question, the how of psi. There are also a host of technical reviews of the experimental literature on parapsychology, known as meta-analyses, which merge most of the available published research studies directed towards a particular topic. For example, Utts (1991) published one such review paper in the prestigious statistics journal Statistical Science, and psychology professor Daryl Bem from Cornell University, along with the late Charles Honorton, published a notable review in Psychological Bulletin of the literature on a type of ESP procedure known as the ganzfeld—where psi is observed when the participant is in a mildly altered state of consciousness (Bem & Honorton, 1994). Furthermore, since the publication of the AIR report on the STAR GATE program 19 years ago (Mumford et al., 1995), scientists have continued evidentiary and explanatory studies using improved methods and technology, the results of which cannot be dismissed outright.
Although doubts have also been raised about the utility of psi for military purposes, it is important to note that many eminent scientists, including academicians and Nobel laureates, have supported these programs. Prominent scientists have been involved in the exploration of these phenomena from the early days of its experimental investigation in the late 19th century. While officials of the American government involved in the STAR GATE program may be reluctant to “come in from the cold,” the most prestigious award—the Legion of Merit—given to McMoneagle for his excellence in providing psychic intelligence in peacetime, is evidence of their support and satisfaction with the applied aspect of the program.

Not only does this lend credence to the existence of effective psychic functioning, it hints at the U.S. government’s high-level involvement, which reached all the way to the West Wing of the White House. Needless to say, this acknowledgement does not support the doubts raised by the reviewers of the program on the value of psychic espionage as one of the methods of intelligence gathering.

I was visiting a senior senator on the Senate Select Committee on Intelligence, who asked if I could rebut the report’s conclusion. I said it would be simple, and he asked that I do so, and, to quote him, “pull no punches.” That published report can be found in May (1996). How did all this affect me personally? One thing is obvious. I learned to be a program manager with a substantial budget and a group of very bright people who often held firm and diverse opinions on nearly everything. But something more important may have happened to me and my worldview. In modern consciousness studies, there is a spread of ideas about the nature of consciousness, ranging from the dualist perspective, on the one hand, that some part of us, for example our soul or some other non-material aspect, survives our death, to the materialist point of view, that mind and brain are the same and death is the end of an individual’s consciousness. More technically, according to the latter view, mind or our rich internal and subjective experience is an outgrowth (a.k.a. an emergent property) of the vast number of neurons in our brain and the even larger number of interconnections among them. This is the view that I have arrived at based on the data, my experiences, and a growing accumulation of supportive research data. Currently this reductionist/materialist view is held by a very small minority of researchers currently active in trying to understand ESP, although it is easily the consensus within the neurosciences and research psychology communities.

One fallout of the Cold War conflict finally being over is that I have joined forces with our former enemies from the Soviet era to jointly author a book on how both sides approached the art of psychic warfare (May, Ruble, & Auerbach, 2014). This has turned out to be a much more difficult task than I had originally thought. Although atheism was part of the official Soviet dogma, on a working level it was simply ignored by many. Even the Russian Orthodox Church has its unofficial roots in Russian shamanism, and the good news is that in shamanistic traditions ESP is considered a good thing. Thus, we had no difficulty at all getting quotes and the support of former very senior officials in the Soviet system, who were quite happy to come forward and admit their interest in ESP.

On the American side of the ocean, however, we are generally a Protestant nation, and in many denominations of that tradition ESP is considered the work of the Devil. We also enjoyed top-level government supporters, which included a Secretary of Defense and other agency directors, most all of whom have retired by now. None of them, however, are willing, as the spy stories say, to “come in from the cold.” They will not allow me to mention their names, even though the evidence of their involvement in the programs is now part of the public record.

I have had the pleasure of visiting Moscow maybe a dozen times by now, and these former “enemies” have become good friends. Thus, during one of my many visits to Moscow, I met with three of my Russian coauthors and our host in his office. Major General Nikolai Sham, Deputy Director of the KGB (Ret.), who kindly wrote the foreword to our book, had also joined us. All of the Russians present had been members of the Communist Party, which officially implied they were firm atheists and materialists. From our discussion of the nature of consciousness, we realized that there was one, and only one, materialist-atheist in the room. Me! The rest were hard-line idealists and theists. We had a great laugh over the obvious irony.

Conclusion

Psychic phenomena have been part of the human experience ever since we, as a species, could communicate, or possibly even before the evolution of language in other species. The skeptic might say, with considerable justification, that much of the putative psychic reporting has been little more than fantasy, selective memory, or some other form of self-delusion. Beginning with the founding of the Society for Psychical Research in London in 1882 and the later pioneering efforts of J. B. Rhine at Duke University, scientists have been challenged to determine
what, if any, of these remarkable self-reports can be teased into the laboratory and studied according to the rules of the scientific method. Perhaps surprisingly, a great deal of research has passed that requirement.

For example, card-guessing experiments, initiated by J. B. Rhine and colleagues at Duke University (Rhine, 1934, 1936), were controversial but taken seriously, culminating with Price’s (1955) review of that research in Science. Price acknowledged that Rhine’s results were revolutionary if true, but because he could not identify any technical or procedural flaws or think of any other viable explanation, he concluded that the results must have been due to fraud. This suggestion spawned a lively debate in Science in the 1950s (Bridgman, 1956; Meehl & Scriven, 1956; Price, 1956; Rhine, 1956a, 1956b; Soal, 1956). Two decades later, Price (1972) apologized for his unfounded suggestion. The present public brouhaha regarding the studies on precognition by Bem (2011) is reminiscent of the controversy created around the Rhine data by Price.

As project leader for the research section of STAR GATE, I subcontracted nearly one million dollars over the years to qualified researchers in the field. In one of those contracts, I asked Charles Honorton to conduct a detailed meta-analysis of the precognitive Zener card guessing data. He and Ferrari conducted a meta-analysis to examine all such experiments published from 1935 to 1987 (Honorton & Ferrari, 1989). Their assessment of all such experiments between 1935 and 1987 showed a cumulative statistical effect 11.4 standard errors over chance expectation: a knock-your-socks-off result. Moreover, they determined that neither selective reporting practices nor variations in study quality could account for the observation that, on average, human subjects were able to correctly guess the symbol of a randomly-determined future stimulus card (slightly but significantly) more often than expected by chance. These results stand to this day.

It is beyond the scope of this article to provide an assessment of all the anomalous cognition (a.k.a. ESP) research spanning the last 80 years or so. But perhaps the best evidence for the existence of anomalous cognition arises not from pure academic pursuits, but rather from successful applications. The intelligence community, for example, could not care less about the mechanisms of anomalous cognition. It certainly worked well enough for them to keep the program alive for two full decades. I realize, of course, that the cynical reader will simply observe that our stupid government kept lots of dumb things funded. Hence the Golden Fleece Awards. The defense I offer against that accusation is simple. There were many people in the government who wanted to shut the program down even from its beginning. If it were not for a handful of heroes who put their considerable weight and reputations on the line supporting our project, these detractors would clearly have been successful.

I am often asked, “Is the government still involved?” As I gave up all my security clearances long ago in the spy versus spy game, I simply cannot say for certain. But in my opinion, it is not funding further work. Given the state of the world just now in 2014, all I can do is hope that my assessment is incorrect.

References


Laboratories for Fundamental Research
Palo Alto, CA, 94306, USA
may@lfr.org

Abstracts in Other Languages

**German**

STAR GATE: DAS PARAPSYCHISCHE SPIONAGEPROGRAMM DER U. S.-REGIERUNG


**Spanish**

STAR GATE: EL PROGRAMA DE ESPIONAJE PSÍQUICO DEL GOBIERNO NORTEAMERICANO

RESUMEN: Desde 1972 hasta 1995 varias agencias del gobierno de EEUUAA subvencionaron aplicaciones e investigación en psi con una suma de más de 20 millones de dólares. Aunque esto suena como una cantidad considerable de dinero para la mayoría de nosotros, para los militares y/o financiación de inteligencia es apenas cambio en el bolsillo! Esta actividad no se inspiró en la curiosidad académica que uno podría encontrar en una universidad; más bien fue impulsado por una necesidad durante la Guerra Fría. Puthoff y Targ merecen un enorme respeto por haber guiado el proyecto, especialmente en sus primeros días. Mi perspectiva de Star Gate se extiende desde finales de 1975 hasta 1995 y yo soy el “guardián” de toda la investigación y parte del espionaje. Esto significa que tengo todos los archivos del proyecto, incluyendo los datos en bruto de un gran número de experimentos, los informes finales a una gran cantidad de clientes, registros administrativos sobre quién financió el proyecto y con cuánto dinero, quién
STAR GATE : LE PROGRAMME D’ESPIONNAGE PSYCHIQUE DU GOUVERNEMENT ETATS-UNIEN

RESUME : De 1972 à 1995, diverses agences du gouvernement états-unien ont financé des recherches et des applications du psi à hauteur de 20 millions de dollars U.S. Bien que cela semble représenter un montant important pour la plupart d’entre nous, pour le budget de l’armée et/ou du renseignement, c’est une somme négligeable ! Cette activité n’était pas inspirée par une curiosité académique similaire à celle que l’on rencontre dans une université ; il s’agissait plutôt d’une nécessité durant la Guerre froide. Puthoff et Targ méritent un grand respect pour avoir chapeauté le projet à son démarrage. Ma perspective sur STAR GATE s’étend de 1975 à 1995, et je suis le « gardien des clefs » de toutes ces recherches et de certains espionnages. Cela signifie que j’ai en ma possession tous les documents de travail, dont les données d’un grand nombre d’expérimentations, les rapports finaux remis aux clients, les documents administratifs pour savoir qui a financé quel projet et à quelle hauteur, qui était impliqué, et comment et pourquoi cette activité interne au gouvernement fut établie à Fort Meade. Cet article est un récit personnel de première main sur une grande partie de ce travail.
THE NATURE OF PRECOGNITION

BY JON TAYLOR

ABSTRACT: This paper describes a theory explaining precognition literally as the “pre-cognition” of information contained within the percipient’s brain in the future—a link with his or her future experience of the event. The theory is based on the block universe model, in which past and future events already exist in the space-time continuum, as required by the special theory of relativity. Bohm’s theory of the implicate order is compatible with such a model, and it suggests that if similar structures are created at different locations in space and time, the structures resonate with a tendency to become more similar to one another. The principles are applied to the neuronal spatiotemporal patterns that are activated in the brain. Precognition is considered to be the fundamental phenomenon of ESP and manifests as information transfer from the brain in the future to the same brain in the present. The model allows also for the possibility of contacts with other brains, and these contacts would occur either in real-time or at different times. However, direct contacts with external objects or events are considered not to occur at all. The mechanism is applied to experiments in precognition, and it explains the apparent anomalies found in the results.

Keywords: precognition, intuition, block universe, implicate order, intervention paradox

Many attempts to explain extrasensory perception have focussed on the phenomenon of precognition, which may be defined as the supposed ability to obtain cognitive information about a future event that could not otherwise be anticipated through any known inferential process. Precognition is often considered more problematical than real-time ESP, because it requires that information somehow has to be transferred backwards in time. Nonetheless, there is considerable evidence for its occurrence.

A meta-analysis published by Honorton and Ferrari (1989) covers all the forced-choice precognition experiments carried out from 1935–1987. The data base includes 309 series of experiments with over 50,000 participants and a total of nearly 2 million trials. A small but reliable effect was found (effect size $r = .01$, Stouffer $Z = 6.02$, $p = 1.1 \times 10^{-9}$). A more recent meta-analysis (Storm, Tressoldi, & Di Risio, 2013) included a further 33 precognition studies carried out from 1987–2010. A comparison of these meta-analyses shows no decrease in effect size over the entire survey period, during which the research quality improved considerably. It suggests that the early results could not have been artifacts of poor design.

Because the targets used in the experiments were not selected until after the participants had made their guesses, the results give outstanding evidence for contacts with something in the future. However, the experiments do not tell us with what the contact is made. Is it direct contact with the inanimate target object or someone else’s mental impression of that target? This is the alternative currently preferred by the majority of parapsychologists. Or is the contact with the participant’s own future knowledge of the target, knowledge obtained when the participant receives feedback of the target information? This is the alternative I am proposing.

Existing theories have already been reviewed (e.g., Carr, 2008; Rush, 1986; Stokes, 2007) and need not be further discussed here. However, several of the theories invoke the influence of a nonphysical consciousness, an interpretation which a majority of mainstream scientists would reject. The present theory adopts the materialist view in which conscious awareness is considered to occur only as an epiphenomenon of the neural processes involved.

The theory explains precognition as information transfer from the brain in the future to the same brain in the present. The theory is based on the block universe model and on David Bohm’s theory of the implicate order. It suggests that if similar structures are created at different places and different times, the structures may be said to “resonate,” with a tendency to become more closely similar to one another. The principles are applied to the

---

1 An earlier version of this paper was presented at the 56th Annual Convention of the Parapsychological Association, Viterbo, Italy, August 8–11, 2013.
neuronal spatiotemporal patterns that are activated in the brain, to show how information transfer could be produced. For example, a pre-cognition would occur if the pattern activated at the time of the future experience of an event resonates with any similar pattern that is (spontaneously) activated in the present. This could enable the level of activation of the present pattern to build up to the threshold at which it produces the conscious awareness of an event similar to the event that will be experienced in the future.

Thus, precognition is explained as a connection with the percipient’s own brain in the future (a link with his or her future experience of the event), and telepathy might similarly be explained as a link with the brain of another person. This simplifies the concept of ESP considerably, because it eliminates the need for direct “clairvoyant” contact with the event itself. Furthermore, the theory explains the results of micro-PK experiments in terms of precognition, so that the theory may be considered to offer an explanation for most classes of psi phenomena.

The present paper clarifies several of the issues raised in my earlier paper (Taylor, 2007), in which I attempted to show that retrieval from the past (ordinary memory) and retrieval from the future (precognition) may occur in much the same way.

The Block Universe Model

Contact with an event in the future suggests that the future event must “already exist” in some sense (Broad, 1978, p. 305). This conforms to the block universe model, in which past and future events already exist in the space-time continuum, in accordance with the special theory of relativity. The model combines the three coordinates of space with one coordinate of time to create a frozen-block universe of four-dimensional Minkowski space-time. Such a universe could provide the information base necessary for precognition to occur (Werth, 1978, p. 54).

People often object to the fatalistic implications of a determined universe. However, a number of physicists and philosophers have argued strongly in its support (e.g., Petkov, 2004; Putnam, 1967; Rietdijk, 1966). Paul Davies (2002) reviews current thinking and suggests that “physicists think of time as being laid out in its entirety—a timescape, analogous to a landscape—with all the past and future events located there together” (p. 42).

There would appear to be little room for free will within such a model. However, in an article in Nature, Kerri Smith (2011) suggests that conscious free will may be just an illusion. The article reviews a number of experiments which show that the outcome of a voluntary decision can be encoded in brain activity up to several seconds before the decision enters conscious awareness (e.g., Fried, Mukamel, & Kreiman, 2011; Libet, Gleason, Wright, & Pearl, 1983; Soon, Brass, Heinze, & Haynes, 2008). Smith points out that some philosophers have objected that the experiments refer only to simple decisions, and a project financed by the Templeton Foundation has been set up to resolve this question. Further arguments against free will are given by neuroscientist Sam Harris (2012), and physicist Jean Burns (1999) discusses the difficulties that would be involved in adapting current physical law to accommodate free will. The consensus is beginning to support Baruch Spinoza’s (1677) conjecture that “Men believe themselves to be free, simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined.”

One apparent difficulty with the block universe model is that it seems to be incompatible with the ability of the brain to make decisions that affect the future. To illustrate this point, we shall look at a simplified model in three dimensions. We omit one of the coordinates of space to show the coordinate of time. Figure 1 (see next page) shows a block with the dimensions 10,000 km x 5 km x 80 years. The scale chosen for each of the coordinates is arbitrary and, for simplicity, the movement and curvature of the Earth are ignored. Space and time are shown as fixed extensions; there is no flow of time within the block itself. According to Einstein, the flow of time is a psychological effect due to the movement of a person’s consciousness through the block.

We now look inside the block and introduce a person, Harry. We suppose that Harry lives for 80 years: He is born at point A (a village in England, say) and 80 years later he dies at the same place, at point A’. We can show Harry’s movements in two dimensions of space, for example, from side-to-side (East-West) as well as up and down, as he travels forwards through time. We’ll suppose that when he is 20 years old Harry climbs a nearby mountain (point X), and when he is 60 he makes another journey, first to the West (point Y) and then to the East (point Z). Finally, Harry dies when he reaches point A’. The zigzag line represents Harry’s worldline through space-time.

We shall assume that Harry’s present lies within a plane, such as PQRS, which represents an infinitesimally thin slice of space-time—an approximation because relativity theory allows simultaneity at different points in space only for observers in the same frame of reference (Burns, 1999; Carr, 2008, p. 60). The plane travels through the
The Nature of Precognition

The Nature of Precognition

block as shown by the arrows; it has in front the events that occur in the future, and it leaves behind the events that have occurred in the past.

Figure 1. The block universe.

Now, if we look from the moment of the “present”, we can see how the events in Harry’s “past” (such as climbing the mountain) are events that resulted from decisions made at that time in his brain. Similarly, the events in his “future” (such as the journey) are events that will result from future decisions. But these events already exist in space-time; they cannot be changed. Harry cannot take a shortcut through time and influence either the past or the future. He can only influence each moment as it becomes the “present,” and the diagram already shows what his brain decides to do at each moment. Thus, as the physicist D. F. Lawden (1999) puts it: “We act as we choose, but what we choose is rigidly determined” (p. 184).

This overcomes the objection that the block universe is not compatible with the possibility of influencing future events. The difficulty seems to arise because we imagine that we can influence a future event by acting in the present. Instead, the influence is produced only at a later moment, when the “future” becomes the “present.”

Notice, however, that if we look from outside the block, we see that Harry doesn’t actually make a decision; instead, he becomes aware of the decisions and decision-making processes that are laid out along the coordinate of time. Each of the decisions is determined by chains of neuronal activation which carry information received from his senses and from the activation of his memories. The major challenge for physicists relates to the flow of time. The contents of the block are fixed, and it is only our conscious experience that changes as the experience moves towards the future (e.g., Penrose, 1995, p. 384; Smythies, 2003, pp. 52–55; Werth, 1978, p. 53). We shall return to this point in the next section.

Interpretations of Quantum Mechanics

The concept of a determined universe does, however, conflict with the orthodox version of quantum mechanics. According to this version, the future outcome of a quantum process is probabilistic, so that it is impossible to know the nature of the outcome until after that outcome has been produced. Some theorists have used this to try to “explain away” true precognition as being due to contacts with what will be probable, rather than actual futures (see Radin, 1988). However, forced-choice target-guessing experiments are normally designed to ensure that targets are selected with equal probability, so that the positive results do not necessarily require an a priori knowledge of options that are more likely to be selected.

Both David Bohm and Yakir Aharonov have proposed that the future outcome appears to be uncertain only because we do not have access to the information necessary to deduce what that outcome will be. For example, Bohm suggests that there is a hidden order at work behind the apparently indeterminate nature of matter as described by the orthodox version (Albert, 1994; Bohm, 1957, 1983). He proposes that behind the
quantum-mechanical processes, which obey statistical laws, there is a deeper level of activity involving subquantum-mechanical entities that are subject to their own laws. Bohm (1951) calls these entities and their laws hidden variables, which some physicists consider controversial. The hidden variables influence the collapse of the wave function and determine which of the potential outcomes (represented by the Schrödinger equation) will become the actual outcome.

Bohm’s theory simplifies quantum mechanics, even though it involves the concept of hidden variables. However, Carr (2008, p. 29) points out that whilst experiments have refuted some models which use hidden variables, they have not refuted Bohm’s model. In fact, Musser (2004) discusses an approach which specifically involves influences coming from the future. He suggests that quantum mechanics seems odd, because we assume that only the past affects the present; but if the future does too, then the probabilistic aspect of quantum mechanics could be explained because we don’t know what will occur in the future. Musser refers to Hadley (1997), who points out that according to relativity theory, past and future already exist; thus, it would be natural for both to affect the present.

Aharonov and colleagues have taken this a step further. In their paper “A Time-Symmetric Formulation of Quantum Mechanics,” Aharonov, Popescu, and Tollaksen (2010) propose that there are two wave functions: one that propagates forwards in time, and another that propagates backwards in time. They suggest that the apparently random events in quantum mechanics are caused by events in the future, and that the wave function coming from the future is what carries the missing information. Furthermore, they successfully carried out what they called “weak measurements” to show that a future cause can lead to a present effect (Aharonov, Albert, & Vaidman, 1988; Aharonov et al., 2010).

The interpretations proposed by Bohm and Aharonov are both deterministic and allow for a backwards transfer of information through time. However, Bohm has developed his ideas to include the wider concept of an implicate order, which is based on the zero-point (vacuum energy) field that extends throughout space and time. It is from within the implicate order that the hidden variables exert their influence on the wave function to create the manifest universe. Bohm suggests that the implicate order constantly unfolds to create the successive instances of the manifest universe—the explicate order—and that these instances are immediately re-enfolded back into the implicate order. The notion of continuity of existence is given by “the rapid recurrence of similar forms” (Bohm, 1995, p.183). It is rather like the frames of a film projected onto the screen in a cinema; the individual frames give rise to a continuous image in which there is a smooth transition from one frame to another.

This can be visualized by referring to the block universe model (Figure 1). The implicate order is spread throughout all space and all time (e.g., throughout the block). It is constantly unfolding, along with consciousness, into the plane PQRS, thus creating the successive slices of space-time. The process of unfoldment and re-enfoldment is therefore considered to be fundamental (Bohm, 1995, p.185; Peat, 1987), and such a process leaves behind the structure that we call the block universe.

Bohm’s theory suggests a way information could be transferred, because it proposes that similar structures resonate within the implicate order and tend to unfold in a form which makes them increasingly similar to one another (Bohm, 1990, p. 93; Bohm & Weber, 1982, p. 36). The term “resonance” here refers to a kind of mutual attraction between the structures, and it should not be confused with the term normally used in physics.

Notice, however, that all structures might resonate to some extent, and not only those which are similar. Nonetheless, the effect of the resonance could be stronger in the case of similar structures. This would overcome Susan Blackmore’s (1985) criticism relating to Rupert Sheldrake’s (2009) morphic resonance on the grounds that morphic resonance leads to a “causation loop”: the structural similarities cause the resonance, and the resonance causes the similarities. The present theory overcomes this problem because the resonance occurs anyway, regardless of the degree of similarity.

Bohm suggests that influences coming from past forms enable the past forms to be replicated or repeated in the present—thus leading to the formation of relatively stable structures. But he doesn’t mention the possibility that future forms might influence present forms. His concern was that the orthodox version of quantum mechanics fails to account for the notion of temporal process and repetition of form, and he introduced the concept of resonance to show how a fixed disposition can be established as the forms are successively unfolded. However, when referring to the implicate order, he says that “Since no space and time are relevant there, all things of a similar nature might get connected together or resonate in the totality” (Bohm, 1990, p. 93).

Therefore, let us suppose that two forms, A and B, are unfolded at different times. If the forms resonate in
the implicate order, then A influences B just as B influences A, regardless of which one unfolds in the future. The forms mutually influence each other, and this influence represents an exchange of information between them.

**The Mechanism of Information Transfer**

The forms that resonate in the implicate order correspond to *processes* in which events cause one another to occur in our manifest universe of space-time events. When activity in the environment is perceived by our senses, the perceptual information is represented by a spatiotemporal pattern of activation in the brain. If similar patterns are activated at different places and different times, the patterns resonate in the implicate order with a tendency to unfold in a form in which they are more similar to one another.

However, it is unnecessary for the entire patterns to be similar. An association between any pair of networks represents a process in which the activation of one network causes the activation of the other. Thus, if the same pair of networks is activated on different occasions, the replication tendency will affect that pair of networks. It will enable a transfer of information. The important point here is that the transfer refers to the *association* between the elements of information represented in the two networks, and not to the discrete elements of information represented in either network on its own. As we shall see later, this is necessary to avoid a confrontation with the intervention paradox.

Notice that the resonance does not produce changes in a pattern after it has been unfolded; the pattern is unfolded in a form in which it has already been affected by the resonance. Thus, for precognition to occur, the brain has to detect the fact that the pattern has unfolded in a form which is slightly different from that in which it would have unfolded if it hadn’t been affected by a resonance. Any transferred information is already inherent in the structure that is unfolded, and there is no *transmission* of information through time.

The theory contrasts with some dualistic theories that involve consciousness as the medium for the information transfer. For example, John Palmer (1995) proposes a theory which is similar in some ways to the present one, as it requires that the brains are in similar states and it rejects clairvoyance. Palmer proposes a two-way interaction between brain and consciousness, which allows them to form a closed system which is also compatible with a determined universe. However, the information transfer takes place via consciousness, whereas the present theory proposes a direct transfer of information between the resonating neural structures.

**Long-Term Memory**

To understand how information is retrieved, either from the past or from the future, we need to consider briefly some aspects of the long-term memory system. Memories about events, about ourselves, and about our previous experiences with the events are stored in neuronal networks. When (internal) stimuli arrive at the networks, chains of impulses travel along various pathways and trace out a highly complex pattern that has extension in space and duration in time—a *spatiotemporal pattern* of neuronal activation. If the pattern is activated repeatedly and the degree of activation reaches a certain threshold, it may give rise to conscious awareness of the information contained in some of the networks (Taylor, 2007, pp. 556–559).

When we experience an event through the senses, stimuli from the sensory areas are applied to the memory networks. According to the representative theory of perception (e.g., Smythies, 2003; Smythies & Ramachandran, 1998) the brain activates an *existing* network that represents the experience. The perceptual stimuli are matched to the network because similar stimuli have contributed to its formation on previous occasions. However, with elements of new experience, new associations are created which are added to the existing networks.

The incoming stimuli activate both the networks corresponding to the event and networks corresponding to one’s previous experiences of similar events. An *appraisal* may be carried out if the event is significant (i.e., beneficial or harmful) to one’s personal well-being.

Neurophysiologist Joseph LeDoux gives as an example the evaluation and appraisal of the experience of “seeing a snake” (see Figure 2). The incoming stimuli activate the networks representing the snake and networks representing one’s previous experience with snakes. Because the experience refers to a potentially harmful event, the outgoing stimuli trigger the appraisal networks in the amygdala. The appraisal leads to an emotional feeling of fear, and it may lead to a physiological response, such as jumping away from the snake. The amygdala is connected, via re-entry circuits, back to the memory networks, and this can lead to many more reactivations of the pattern produced (LeDoux, 1993, 1998, pp. 283–296).
When a memory is retrieved, the neural impulses follow similar pathways and activate a pattern that is similar to the pattern activated at the time of the initial experience. The fundamental question is how do the impulses "know" which pathways to take through the networks on retrieval? My theory suggests that various patterns are activated until one of them matches and resonates with the pattern activated at the time of the initial experience (Taylor, 2007, pp. 559–562). Such an experience can take place either in the past or in the future, so that past and future memories may be retrieved in a similar way. C. D. Broad (1978, pp. 289–296) discusses the close relationship that exists between memory and precognition when considered from a philosophical point of view. In the present paper, I am concerned specifically with the precognition of future events, and I shall classify the events according to whether they are caused by outside circumstances, or by the percipients themselves.

**Future Events Caused by Outside Circumstances**

Let us consider a simple case of precognition. We will suppose that a percipient, Tom, experiences an emotional event when it occurs tomorrow (see Figure 3). During tomorrow’s experience, he activates a pattern which includes networks representing the event and networks representing his previous experiences with similar events. The pattern is activated repeatedly, and re-entrant stimuli from the appraisal networks lead to further reactivations. Apart from consolidating the memories being formed, these repeated activations contribute to the formation of a stronger field which, in turn, influences any similar pattern that is independently activated elsewhere in space and time.

In fact, Bohm suggests that “formative fields” may be associated with the structures in the implicate order, and that the fields resonate and produce a tendency towards replication between the structures as they are unfolded. For example, when referring to Rupert Sheldrake’s morphic resonance, Bohm (1990) says that “You could picture a formative or morphic field as a very subtle aspect of the implicate order that would impress itself on the denser and explicate energies” (p. 99). Sheldrake (2009, 2012) uses these principles in his theory of morphic resonance, in which he suggests that past forms tend to be replicated or repeated in the present.

Thus, for Tom to precognize the event today, the neural impulses may have to follow various pathways through the networks until a pattern is activated that is similar to the one which will be activated tomorrow. A resonance is produced, and according to Bohm’s theory, the synapses forming today’s pattern unfold in a form in which they are more similar to those forming tomorrow’s pattern.

However, when tomorrow’s pattern is activated, the continuing flow of impulses through the networks causes the synapses to be strengthened (over a period of time) and this facilitates the flow of further impulses. The weaker synapses forming today’s pattern are also strengthened (in this case due to the resonance) and this also facilitates the flow of impulses. It may result in the flow of a few extra impulses, which lead to stronger output stimuli being produced. These stimuli go back to the networks via re-entry circuits that go through the working memory system. If stronger stimuli are applied to the working memory cells, these cells can sustain their activation, and this may overcome the effects of inhibition and lead to a sustained activation of the networks concerned (Fuster, 2003, pp. 155–164; Taylor, 2007, pp.558–559). A precognition occurs if the threshold is reached at which the activation of
The Nature of Precognition

The networks is sufficient for Tom to become consciously aware of the information represented by these networks. Tom is actually recalling today the “memory” of an event that he will experience tomorrow.

![Figure 3. Precognition.](chart)

Notice that Tom associates the event with his own experiences of it, both in the present and in the future. This enables selectivity, because the precognition will refer to the event experienced by Tom, not to a similar event experienced by someone else. The association between the event and Tom’s experiences of it may be described as an associational link, which connects the present and future patterns, and ensures a stronger resonance.

In fact, telepathic contacts with other people may be far more difficult to produce. For example, suppose that a sender, Mary, experiences an emotional event. Her brain activates a pattern containing a representation of the event, a representation of her previous experiences with similar events, and a representation of Mary herself (i.e., her body image). Meanwhile, Tom happens to activate a pattern corresponding to a similar event. In his case, the pattern contains a representation of the event, a representation of Tom’s previous experiences of similar events, and a representation of Tom himself. The patterns do not match and any resonance between them will be much weaker. For this reason, many ostensible telepathic contacts with other people could be explained as pre-cognition of the percipient’s own future knowledge of the event—knowledge he obtains when he is later informed about it. In some of the early telepathy experiments (e.g., J. B. Rhine, Pratt, Stuart, B. Smith, & Greenwood, 1966) it was found that removing the sender did not affect the results. This suggests that telepathic contacts with the sender were not being detected in these experiments.

**Future Events Caused by the Percipient**

Sometimes, a person will do something to cause (or allow) an event to occur. The brain activates a pattern containing the association between “something the person does” and “the effect that it produces.” Again, we refer to a process in which a cause produces an effect.

**Intention Fulfilled**

Suppose that our percipient, Tom, is thinking about taking a train to London. He activates networks in his brain through which he associates the idea “taking the train” with the idea “arriving in London.” As the degree of activation increases, it may lead to an intention being created to carry out the corresponding action (see Figure 4). Further activation leads to a triggering of the motor networks, so that Tom goes ahead with the intended action. We will suppose that Tom later fulfils the intention, so his future experience is that of taking the train and arriving in London. Because the present and future associations are similar, a resonance occurs, and Tom might be able to precognize the knowledge that he will be successful in fulfilling the intention.

However, the precognition occurs only if Tom does take the train, and it does enable him to arrive in London. The resonance is between the patterns containing the association between these two elements. For example, if Tom were to precognize his arrival only, he could then decide not to take the train and he would confront the intervention paradox. This is why I emphasized that the resonance always refers to a complete process, including the cause and the effect, even if the activation of one of the networks does not reach the threshold at which it produces a conscious awareness of the information represented by that network.
Intention Not Fulfilled

Again, we’ll suppose that Tom intends to take the train to London, but on this occasion the train is destined to crash. Whilst Tom is intending to carry out the action, there is no resonance, because there will be no future experience of taking the train and arriving in London. If Tom does take the train, it will lead to him experiencing the crash (see Figure 5). The intention is therefore not fulfilled and a precognition does not occur. This detection of the absence of an intended future could form the basis of an intuition, in which Tom might become aware of a possible danger, perhaps without knowing exactly what the danger is. Thus, the absence of the precognition would serve as an intuitive warning which tells Tom that something will prevent him from fulfilling the intention. It doesn’t tell him specifically that it would be due to an accident, but it might result in a feeling that “something is wrong.” This enables Tom to change his intention and decide to do something else, such as staying at home.

The mechanism shows how the intervention paradox is avoided. If a percipient were to precognize an event-causing process (i.e., “scenario”) that he intended to change, and then he did change the scenario (i.e., created a different scenario) he would confront the paradox. For example, in the situation depicted in Figure 5, Tom’s initial intention is to take the train to London. Later, Tom does take the train and he experiences the crash. There is no resonance between the brain patterns representing the present intention and future experience because their content is different, so Tom cannot obtain any information about the crash precognitively. This was originally stated as the principle of intentionality (Taylor, 1995), which says that you cannot obtain a precognition about a given scenario, if the scenario is one that you intended to change.

Evidence that people can use intuition to avoid accidents is given by W. E. Cox (1956), who carried out a survey of passenger statistics on the U. S. rail networks. He found that significantly fewer people travelled on trains involved in accidents than on comparable trains not involved in accidents. The survey doesn’t tell us what made the passengers stay off these trains, but many case histories suggest that people simply find themselves looking for...
something else to do, without actually being aware of the potential danger that lies ahead (Dossey, 2010, pp. 30–44).

We therefore conclude that the block universe is built up in a fully determined form, and it cannot subsequently be affected by intervention. Otherwise it wouldn’t be a block universe.

**Cases of Apparent Intervention**

Cases have been reported in which it does appear that the percipient has successfully intervened to prevent a precognized event from occurring (e.g., L. E. Rhine, 1955).

**Future Events Caused by Outside Circumstances**

A person precognizes an event and then uses the information to change another event as a result of knowing that the precognized event will occur. An example often quoted concerns a woman who dreamed of a chandelier falling in her baby’s bedroom and killing the baby (Rhine, 1961; Stokes, 2007, pp. 84–86). As a result, the woman collected the baby from the bedroom. Later the chandelier did fall, but without harming the baby. This appears to be an act of intervention.

However, the precognition itself refers only to the falling chandelier. This event was destined to occur, regardless of whether or not the woman had acted to save the baby. The woman therefore deduced that there could have been a danger to the baby, and in the way typical of dreams (e.g., J. W. Dunne, 1981), the fantasy about the baby being killed found its way into the dream narrative.

**Future Events Caused by the Percipient**

This kind of situation requires intuition. A person appears to intervene, but really detects the absence of a precognition referring to the outcome of an intention. This serves as an intuitive warning that something is wrong. Take as an example passengers who feel uneasy about taking a train and decide to stay at home. Later, upon reading the news of an accident, they assume that they must have precognized their involvement in the accident, and then they intervened to avoid it.

However, the only information they obtained was the knowledge that they would be unable to fulfil an intention to do something later in the day, and that enabled them (unconsciously) to deduce the possibility of the accident happening. Thus, the percipient never intervenes to change the event precognized and the mechanism can be applied to all the cases of apparent intervention.

**Laboratory Experiments**

**Future Events Caused by Outside Circumstances**

Perhaps the best examples in this category are the pre-stimulus response experiments, in which changes in a given parameter, such as the participant’s electrodermal activity, are produced prior to the delivery of an emotionally arousing stimulus. Radin (1997) performed an experiment in which a series of randomly selected images were presented to participants at intervals on a computer monitor. Highly arousing images were interspersed with emotionally neutral control images. Increased arousal was observed up to 3 sec prior to exposure to the “emotional” stimulus, but it did not occur during the same period prior to exposure to the “calm” stimulus.

The results suggest that the degree of activation of the neuronal pattern corresponding to the arousing stimulus was sufficient for it to influence similar patterns spontaneously activated prior to that stimulus. The activations of the earlier patterns were sustained to a level at which they triggered the earlier responses.

Several replications of this kind of experiment have been carried out (e.g., Bierman & Scholte, 2002; Spottiswoode & May, 2003; Tressoldi, Martinnelli, Massaccesi, & Sartori, 2009). A meta-analysis of 26 studies, described in a paper by Mossbridge, Tressoldi, and Utts (2012), shows an overall small but significant effect (ES = 0.21, combined $z = 5.3, p = 5.7 \times 10^{-8}$). The authors also report that some of the recent experiments have shown that an increase in arousal can occur prior to exposure to an apparently calm stimulus. However, such a stimulus will still lead to an appraisal if it is personally significant to the individual, regardless of whether it is considered significant.
to others (Lazarus, 1994). For example, a picture of a red rose might, for a given individual, evoke associations that are almost as significant as those evoked by an erotic photograph.

Notice that an early conscious awareness of the stimulus does not seem to be produced in these experiments. Neuroscientist Joaquín Fuster (2003) suggests that there are probably two thresholds: “a threshold for the processing of information and a threshold for the conscious awareness of it” (p. 255). This suggests that a stronger stimulus may be required to elicit the conscious awareness.

In a landmark neurosurgical experiment, Libet, Wright, Feinstein, and Pearl (1979) measured the time required to produce conscious awareness in each of two conditions: first, when a stimulus was applied to a participant’s hand (a pin-prick), and second, when a stimulus was applied directly to the area of the cortex involved in receiving the sensory signals from the skin. The results showed that stimulation to the cortex required approximately 0.5 sec of neuronal activity before awareness was produced. However, when the hand was stimulated, the conscious feeling of “pain” was produced at the same moment that the stimulus was applied. The results were attributed to an “antedating” of the conscious experience (see also Popper & Eccles, 1983, pp. 253–261, 364; Penrose, 1995, pp. 386–387).

The mechanism I have suggested above would explain these results. In the absence of an external stimulus, the receptor organs in the skin spontaneously fire impulses which travel along pathways to the brain. Most of the weaker excitations are filtered out en route, but a few survive to activate perceptual patterns corresponding to the impulses. Normally, these activations would not lead to conscious awareness. However, when the external stimulus is applied, the degree of activation of the corresponding perceptual pattern is increased to the level at which it influences the same pattern when it is activated spontaneously just before and just after the application of the stimulus. The earlier activations are thus sustained, and appraisal leads to the earlier feeling of pain. The conscious pre-perception of the stimulus therefore arises from the sustained activation of an earlier pattern, and not from a subjective referral of the conscious experience. In fact, the earlier sustained activations would help to speed up the processing of sensory information in the brain.

Notice that such a pre-perception is not produced when the cortex is stimulated directly. Eccles says that the perceptual experience is different, and that a “neuronal shock wave” is produced, which has “little resemblance to the pattern of neuronal activation generated by the natural input from the receptor organs” (Popper & Eccles, 1983, pp. 255–256). As the patterns do not match, there is no resonance and nothing to “antedate” the experience.

**Future Events Caused by the Participant**

The best examples here are the forced-choice target-guessing experiments in precognition. The participant has to guess which of a given set of options corresponds to the target that will be randomly selected in the future. We will suppose that feedback is given of the result of each trial, and that the symbol “X” is the target for a given trial.

The future experience of looking at X would seem unlikely to have much emotional impact. However, if symbol X can be associated with the idea of scoring a hit, then the association does have impact. Therefore, I suggest that the precognition refers to an association between the ideas of selecting X and scoring a hit—an association that is then appraised to produce a feeling of satisfaction.

The participant starts by intending for the selection of any one of the options to be associated with scoring a hit. However, if an incorrect option is chosen, there will be no future experience of scoring a hit. The patterns of activation corresponding to the present intention and future experience do not match and there is no resonance. An intuitive warning is produced.

The intuitive warning stops the participant from selecting that option (Taylor, 2007, pp. 558–559, 565), and he unconsciously moves on to another option and repeats the process. When he comes to the correct option, a resonance occurs and he is able to go ahead with his intention to select it. The present intention is subsequently fulfilled, as the future experience will be caused by the association between the selected option and the idea of scoring a hit, when feedback of the target information is given.

The correct option is therefore found by looking for the intuitive warnings, which enable the participant to eliminate the incorrect options. It also prevents a causal loop: a resonance causing the participant to select the correct option and selecting the correct option causing the resonance.

In neurological terms, the brain applies stimulation to the pathways corresponding to each of the target options. If an incorrect option is selected, the pathway will not be reactivated in the future and a resonance is not produced. Inhibition from working memory may then cause the activation of the pathway to cease, thus preventing
the participant from registering the option. However, if the correct option is selected, a similar pathway will be activated in the future and a resonance is produced. The re-entry of stimulation to the working memory system prevents inhibition, and this allows activation of the pathway to be sustained until the motor networks are triggered and the participant registers the option (Taylor, 2007, p. 565).

Notice that the activation of each pathway must be allowed to build until it is close to the threshold at which the motor networks would be triggered. Only at this critical level will the inhibition be effective in preventing the activation of the motor networks in the case of an incorrect option. This suggests that participants must be in an appropriate subjective state for them to score hits. In fact, Rhea White (1964) suggests the subjective conditions necessary, and the present theory explains why these conditions should be favourable (Taylor, 2007, p. 565).

In the case of a forced-choice experiment, the best results are to be expected when feedback is given immediately after each trial (e.g., by allowing the participant to see the target). In the scenario in which he is successful, the participant is immediately able to recognize that his guess has produced the effect of scoring a hit. He is able to create the corresponding association in his brain in the future, and it is easy for this association to be identified with the earlier association created at the moment of the guess.

Suppose, on the other hand, that the participant is only given feedback of the result of a series of trials. Again, we take a “success” scenario in which a positive score is produced. For each trial, the participant tries to guess which of the response options will contribute towards producing the positive score. Then, at the moment of feedback, we’ll suppose that he is able to remember, say, two of the trials in the series. Again he associates the chosen options for each of these trials with the idea of the positive score produced. However, in this case, it is more difficult for the future association to be related specifically to the two earlier associations. Furthermore, the “forgotten” trials in the series are void; the guesses are made by chance, and these trials do not contribute to the positive score produced. The significance of the final result is therefore reduced, because the participant is able to recall only a limited number of trials.

A similar mechanism would apply to free-response experiments, including remote viewing. For example, a participant tries to imagine some of the features that might form part of a remote scene. The participant then makes an intuitive decision to identify a combination of these features which, together, can be associated with the idea of producing a successful result. Later, the participant is taken to the location, sees the same features, and does associate them with the idea of the successful result produced. A resonance is thus produced between the present and future patterns of activation.

Notice that if no feedback is given in either kind of experiment, there will be no future experiences of the kind necessary for intuitive decisions to be made. The results will therefore be expected to show little or no significance.

Notice also that the mechanism for any intuitive decision is difficult to test because participants would not know whether the ostensibly avoided outcome would have occurred had they not acted on the intuitive warning. However, the positive results of the target-guessing experiments may be considered indicative of such a mechanism.

The Nature of Precognition

A number of observations can now be made regarding the nature of precognition. They may help explain the apparent anomalies found in the results of the experiments.

Pre-Cognition Is the Fundamental Manifestation of ESP

For many years, parapsychologists have debated whether the results of “telepathy” experiments are due to contacts with the sender or contacts made directly with the targets (Carr, 2008, p. 24; Stokes, 2007, pp.72–73). However, another interpretation is that the results could be due to precognitive contacts with the participant’s own brain in the future (i.e., with his future knowledge of the target information). Misinterpretation of the nature of the contact could have led to problems of repeatability in some of these experiments.

For example, if there is to be contact with the participant’s future knowledge, feedback must be given to the participant in the future, as discussed in the previous section. But this was not always done. If feedback was given, the participant could connect with his future experience of receiving it, and a positive result could be produced. If feedback was not given, the participant could not use pre-cognition, and the results would be expected to fall either to chance, or very close to chance.
Support to this conclusion is given by the Honorton and Ferrari (1989) meta-analysis, which includes a subset of experiments in which details were provided about the amount of feedback given to participants. The largest effects were shown in the studies in which trial-by-trial feedback was given, and the effect size was reduced when only run-score feedback was given. However, the authors state that none of the studies without subject feedback were statistically significant.

In some well-known telepathy experiments, such as the series conducted at the Maimonides dream laboratory (Ullman, Krippner, & Vaughan, 2002) and the ganzfeld studies (e.g., Bem & Honorton, 1994), highly significant results were obtained when feedback was given on a trial-by-trial basis.

In the remote-viewing experiments at the Stanford Research Institute (e.g., Puthoff & Targ, 1976), feedback was given by taking the participants to the target locations after the trials had been conducted. In fact, the CIA-sponsored Star Gate program (e.g., Puthoff, 1996; Targ, 1996) produced some outstanding results in field trials in which feedback was given. However, this particular technique was rarely used for intelligence gathering (May, 1996; Mumford, Rose, & Goslin, 1995), and this may have been largely due to the impracticability of taking participants to the target locations to receive feedback.

Target-guessing experiments have been carried out in which only run-score feedback was given (e.g., J. B. Rhine, 1969). Here, a salience effect was sometimes observed, in which a greater departure from chance expectation was found to occur at the beginning and end of a run. In this case, the target symbols corresponding to the first and last trials tend to stand out, with a better chance of being remembered in the future. Thus, intuitive decisions could be made to identify the symbols which would be associated with the idea of obtaining a successful result when feedback of the run score was given.

Occasionally, it has been claimed that significant results can be produced without any feedback of the target information. For example, May, Lantz, and Pantaneda (1996) performed a remote-viewing experiment in which feedback was displayed tachiscopically to participants. They found that varying the feedback intensity made no difference to the quality of what May calls “anomalous cognition” (AC). However, May assumed that if feedback is necessary, it is the feedback information per se that produces the AC. If so, an increase in feedback intensity should produce an increase in AC quality. In contrast, the present theory suggests that for an intuitive decision to be made, the feedback must produce a cognitive understanding in the participant’s brain. May points out that the feedback intensities could have been too low for detection, and that would certainly have prevented the participants from cognitively evaluating the targets. Nonetheless, two of May’s participants did produce positive results. This may have been due to them associating some of the imagery they created during the trials with the idea of obtaining a successful result. An intuitive decision could be made to find the imagery they would later associate with the successful result, when feedback of the result was received after the experiment.

In fact, as May himself points out, it is not possible to guarantee that feedback has been completely eliminated. For an experiment to be valid there must be a result, and however well the result is concealed, there is always the possibility that the participant will obtain knowledge of that result at some time in the future.

Interestingly, Steinkamp, Milton, and Morris (1998) performed a meta-analysis of forced-choice experiments comparing clairvoyance and precognition tests, finding the magnitude of the effect to be significant at approximately the same level in both. The assumption was made that direct contact was being made with the targets in each of the two modes, and the feedback issue was not considered. It is therefore possible that, in those studies where there was feedback to the participants, the significant results were really due to the participants precognizing their future knowledge of the target information, regardless of the mode. This suggests that there was essentially no difference between the modes, and that’s what the results showed. There may have been no direct contact with the targets at all.

Notice that when cognitive feedback is given to participants, it enables them to receive and encode the target information using their normal senses. Selectivity is achieved, because the feedback information refers specifically to the target, and the participants do not have the task of trying to distinguish between the target and the decoys.

It would therefore seem important for feedback to be carefully controlled in future experiments and for details of the feedback to be specified clearly in the reports (Taylor, 2008). For example, an experiment could be designed in which trial-by-trial feedback is given to participants in a test group but not to participants in a control group. The future targets are generated for the individual trials in both groups, giving all participants the same opportunity to “clairvoyantly” detect the targets (if such a phenomenon exists). All participants would equally be subject to the possibility of a telepathic influence from the experimenter. However, if there is a difference between
the results from the two groups favouring the test group, it can be attributed to the participants in the test group using precognition to access their future knowledge of the targets. The results of such an experiment could give strong support to the present theory.

**Pre-Cognition Is More Likely to Occur When the Experience of the Event in the Future Produces a Stronger Emotional Impact**

If the emotional impact is stronger, the degree of activation of the appraisal networks is increased, and the re-entry of output stimuli from the amygdala back to the networks will produce many more reactivations of the neuronal pattern involved. The reactivations may be said to produce a *concentration effect* by increasing the strength of the resonance, so that the resonance will be more likely to influence any spontaneous activations of the pattern in the present.

In the pre-stimulus response experiments discussed earlier, it was found that exposure to an emotional stimulus produces a stronger pre-response than exposure to an apparently calm stimulus. In the case of the target-guessing experiments, I suggested that the emotional feeling of satisfaction must be produced when the participants score hits or produce a successful result. For example, if they become bored with the experiment, the association between the selected target symbol and the idea of scoring a hit is no longer appraised in the future to produce the same level of satisfaction. Fewer activations of the pattern are produced, and the strength of the resonance is reduced to a level at which it is less likely to affect the pattern activated in the present. The results should therefore fall closer to chance expectation. This explains the well-known *decline effect*, in which the results do fall closer to chance expectation with repeated testing (e.g., Tart, 1966).

Notice that whilst a high degree of activation is necessary in the future, the degree of activation in the present should be as low as possible, in order for the brain to notice any additional activation which may be caused by the resonance. Excessive activation in the present (e.g., because the percipient is already thinking about the possibility of such an event occurring) will lead to a *dilution effect*, which will reduce the likelihood of the percipient noticing the precognition.

Finally, it has been recognized that extroverts tend to produce higher scores than introverts (Bem, 2011; Honorton, Ferrari, & Bem, 1998) and this is believed to be related to the fact that extroverts are susceptible to boredom and tend to seek out stimulation (Eysenck, 1966). According to my theory, this means that extroverts would actually be expected to perform less effectively in a target-guessing experiment. However, neuroscientist Susan Greenfield (2001) suggests that extroverts (and children) activate smaller neuronal patterns, because the evaluations they carry out of their previous experiences are less detailed—and this leads to stronger emotions being produced. Because the pathways through the networks are shorter, less time is required for the impulses to traverse the pathways and for the re-entry circuits to initiate further activation. This increases the rate of activation of the pathways, and it will produce a stronger emotion and stronger concentration effect, thus increasing the likelihood of extrasensory contact.

**Pre-Cognition Is More Likely to Occur When the Temporal Distance to the Moment of the Future Experience Is Shorter**

To obtain a strong resonance between the present and future patterns of activation, the theory suggests that the synapses forming the patterns have to match one another fairly closely. However, the structures of synapses change constantly over time due to brain plasticity, so that a closer matching is to be expected when the temporal distance between the activations is shorter.

This conclusion is supported by the results of the pre-stimulus response experiments. In Radin’s (1997) experiment, the physiological response was produced 3 sec prior to the stimulus, and in some of the more recent experiments, the response was detected up to 10 sec prior to the stimulus (Mossbridge et al., 2012). On the other hand, the conscious pre-perception of a stimulus seems to be limited to a period of about half a second, which only compensates for the time required to process the stimulus physiologically. In the target-guessing experiments, however, significant results can be obtained over much longer temporal distances. For example, the Honorton and Ferrari (1989) meta-analysis shows that the results were highly significant for feedback delays of up to a few hundred milliseconds. For longer intervals there was a regular decline, due to a weakening of the resonance, and the results
became nonsignificant only when the delays were increased to more than one month. In a study of spontaneous precognition in dreams, Nancy Sondow (1988) found that the number of events precognized declined with the passage of time. Half of the events precognized occurred within one day of the dream, and there was a steep and regular falloff in the number of events with increasing time intervals after the dream.

The limitations imposed by longer feedback delays are important for achieving selectivity in precognition. They enable us to access events in the near future and exclude any similar events that occur in the more distant future.

Once the necessity of feedback has been established, target-guessing experiments can easily be designed to test the effect of varying the feedback delay.

The Results of an Experiment Tend to Conform to Participants’ Beliefs

This is the well-known sheep-goat effect, in which believers in ESP (“sheep”) score significantly above chance, whereas nonbelievers (“goats”) score significantly below chance (Lawrence, 1993; Schmeidler, 1945). To nonbelievers, misses are really hits. They make intuitive decisions to select the target options which, in the future, they will associate with the idea of scoring misses, when they are given feedback of the target information. These are the associations that are appraised to produce the feeling of satisfaction. The participants therefore score a higher proportion of misses, so that the proportion of hits falls to below chance-expectation.

Notice that Schmeidler defines sheep as people who believe that ESP will occur in the experiment. People who believe in the existence of ESP in the abstract sense, but who do not believe that it can be demonstrated experimentally, will therefore produce below-chance scores (Palmer, 1986, pp. 200–204). In fact, the sheep-goat effect is one that relatively few investigators seem to take into account when selecting participants. It could also be responsible for the problems of repeatability encountered when different laboratories perform the same experiment with different groups of participants.

Daryl Bem (2011) performed a series of experiments in which well-known psychological effects were time-reversed. Two experiments were designed to detect a retroactive facilitation of recall, in which participants found that it was easier to recall in the present words that they would rehearse in the future. First, a participant was asked to recall as many words as possible from a list of “stimulus words.” The computer then randomly selected half the words from the list to serve as “practice words” and the participant rehearsed the practice words by performing exercises on them, such as typing them into boxes on the screen.

It is possible that the participants could have detected their future experiences of rehearsing the words. However, such experiences would have little emotional impact, and the number of reactivations of the neuronal pattern would be minimal, occurring only when the participants deliberately reread or retyped the words. Perhaps it is better to interpret the study as if it were a forced-choice target-guessing experiment, in which the stimulus words were the target options and the practice words were the actual targets. The participants could therefore make intuitive decisions to identify the words that they would later associate with the idea of scoring “hits” when they saw these words in the practice list. Although the participants were not actually told that the practice words were targets, they would still obtain emotional satisfaction when they recognized the words they remembered from the practice list. The increased number of reactivations resulting from the appraisals would lead to a concentration effect, thus increasing the likelihood of extrasensory contact.

The participants were, however, told that the experiment was to test for ESP, so that the positive results would depend on their beliefs towards obtaining such results. Participants with a neutral attitude would not appraise the idea of scoring a hit as producing satisfaction; there would be no concentration effect and a nonsignificant result would be expected. Sceptics, on the other hand, would derive satisfaction from scoring misses, and they would tend to produce significant results in the opposite direction; they would recall fewer practice words than control words.

This may have occurred when Ritchie, Wiseman, and French (2012) separately tried to replicate one of these experiments. All three investigators are well known for their scepticism towards ESP, and even if their personal interaction with participants was minimal, the laboratory environment could still have conveyed negative expectations to the participants. Two of the experiments failed to produce significant results in either direction, and the third experiment produced a significant result in the opposite direction.

However, far greater success in replication attempts has been obtained in Bem’s first experiment, designed for the precognitive detection of erotic stimuli. Participants had to guess which of two curtains displayed on a
computer screen hid an erotic photograph. After the participant's selection had been registered, the computer randomly assigned the photograph to one of the curtains. Following a correct guess, the curtain opened to reveal the photograph, and following an incorrect guess, the display showed a blank wall.

A recent meta-analysis (Bem, Tressoldi, Rabeyron, & Duggan, 2014) shows nonsignificant results for 27 replications of the retroactive facilitation of recall experiment, and highly significant results for the 14 replications of the precognitive detection of erotic stimuli experiment (ES = 0.14, combined $z = 4.22, p = 1.2 \times 10^{-5}$). The authors attribute these successful results mainly to the “fast-thinking” protocols used in the experiment which, according to Carpenter (2012), prevent conscious cognitive evaluations from interfering with the intuitive nature of psi functioning. A quick judgement by the participant would certainly be expected to minimize any dilution effect, but I suggest that the high emotional impact produced by the future experience, as well as the very short temporal distance to the moment of that experience, are probably the main factors responsible for the experiment’s success.

Finally, the displacement effect (Crandall, 1991) may be partly attributable to the sheep-goat effect. Displacement occurs, for example, when instead of guessing the target for the trial being performed, a participant guesses the target for the next trial in the series. However, this could easily occur by chance over a number of trials, and it would induce a belief that such a displacement is a normal characteristic of ESP, assuming the participant receives trial-by-trial feedback. Participants therefore assume that they have to guess the displaced targets, and they feel they are scoring hits when they do so. These principles can be applied to any experiment in which participants believe they will select a given response option, regardless of whether that option is the target (e.g., Radin, 1988, discussed earlier).

The Results of REG Experiments May Be Due to Precognition and Not PK

Robert Jahn and Brenda Dunne carried out a program of experiments at the Princeton Engineering Anomalies Research (PEAR) laboratory to investigate the possible effects of consciousness on random event generators (REG’s). The devices were used to emit data streams of 200 binary digits (zeros and ones) for each trial. Participants had to try to influence them, according to pre-stated intentions, to produce high, low, and undeviated mean values for the output distributions. The experiments produced highly significant results (Jahn & B. J. Dunne, 1988, 2005; Jahn, B. J. Dunne, Nelson, Dobyns, & Bradish, 1997).

Edwin May proposed decision augmentation theory (DAT) to explain the supposed ability to influence REG outputs (May, Utts, & Spottiswoode, 1995a, 1995b). He gave evidence to suggest that rather than a micro-PK force acting on the device, the correspondence of the output with the person’s intention depends on optimum sampling of the bit stream generated by the REG.

The DAT model can be explained in terms of intuitive decisions to identify the best moments at which to press the button and initiate the sampling of the bit stream. Suppose, for example, that a participant is trying to obtain a successful result by producing a higher proportion of ones. We assume immediate feedback, and consider two scenarios: In scenario A (failure), at a certain moment, the participant “hears a sound,” intending for it to be associated with the idea of producing a successful result. However, if the participant presses the button at this moment, we suppose there will be no future experience of this occurring—the bit stream will not contain a higher proportion of ones. The present and future patterns of activation do not match and a resonance does not occur. An intuitive warning prevents the participant from pressing the button.

In scenario B (success), at another moment, the participant “notices something” and again intends for what he notices to be associated with producing a successful result. If the participant presses the button at this moment, we suppose there will be no future experience of this occurring—the bit stream will contain a higher proportion of ones. The participant therefore presses the button and, upon receiving feedback in the future, obtains satisfaction from discovering that what was noticed can now be associated with the successful result produced. Resonance occurs between the present and future patterns of activation. The participant has thus been able to make an intuitive decision to choose scenario B, which leads to a positive score.

Notice that scenarios A and B represent response options, in the same way that the features of a remote location represent response options in a remote viewing experiment. Scenarios A and B are temporally separated, whereas the features of a location are spatially separated. But the mechanism is the same.

Notice also that a successful result doesn’t depend on the participant’s physiological control necessary to discriminate one out of a succession of appropriate moments at which the response can be made (May et al., 1995b;
Walker, 1987). Instead, the participant looks for a *scenario* in which the response happens to coincide with one of those moments, and leads to an appropriate sampling of the bit stream. In other words, the task is not that of “looking for” an appropriate moment in time. Instead, it is that of “selecting the scenario which contains a button press that happens to occur at an appropriate moment in time”.

According to this model, the participant’s intention towards the result may be irrelevant. It is the intuitive decision made by the *operator* who initiates the data-selection process which determines the result. This explains the significant results produced when ostensible participants apply the intention at remote locations, or at times before and after the bit streams are generated (B. J. Dunne & Jahn, 1992). The participants themselves do not affect the result; it is the operator who is performing the role of the participant by pressing the button himself. The mechanism also explains why significant results are produced when pseudo-REG’s are used (Jahn & Dunne, 2005). Because the bit sequence is predetermined by a mathematical algorithm, it is more difficult to see how the sequence could be influenced. However, the operator can again make an intuitive decision to select an appropriate sample of the data stream to produce the desired result.

Finally, the results of the series position effect experiments (B. J. Dunne, Dobyns, Jahn, & Nelson, 1994) may have illustrated the effect of the operator’s changing subjective approach as each successive experimental series was performed. The operators showed tendencies to produce better scores in the first series, which fell in the second and third series, and then recovered to an intermediate level in the fourth and fifth series. Initially, the operators would find it easy to adopt the subjective conditions necessary to produce a satisfactory result (Taylor, 2007, p. 565). However, this leads to over optimism, so they either fail to notice the intuitions, or they press the button without waiting for the intuitions to tell them the best moments to do so. The rate of scoring hits is reduced and they become disillusioned; their belief, and the results, now go in the opposite direction. Eventually, a decline effect takes over, and the significance is reduced due to boredom with the experiment.

When the results of these experiments are interpreted in terms of the present theory, they give support to DAT, whilst not supporting the observational theory (Houtkooper, 2002; Schmidt, 1975; Walker, 1974), according to which the observer’s consciousness is supposed to exert a PK influence and bias the collapse of the wave-function. Observational theory has been used to support the clairvoyant interpretation of ESP experiments because it suggests that the participant’s consciousness might be able to collapse the wave-function to create the target state reported.

Bohm’s theory overcomes the observation problem by suggesting that the implicate order still unfolds, regardless as to whether there are observers, or, if there are, whether they are performing an act of observation (i.e., measurement) at the moment of unfoldment. His theory enables us to explain the results of the experiments without having to invoke an influence from consciousness.

**Conclusion**

The theory suggests that psi is manifested principally as pre-cognition of the percipient’s own experience of an event in the future—a process in which information is transferred from the brain in the future to the same brain in the present. A transfer of information from a different brain is also considered possible, but is expected to be far more difficult to detect. According to the theory, the necessity of direct contacts with external objects or events is eliminated. Some of the predictions of the theory are subject to testing, and taking them into consideration may lead to improved repeatability in future experiments.

The theory also suggests that “memories” are retrieved from the future in very much the same way they are retrieved from the past, as discussed in my earlier paper (Taylor, 2007). In either case, the information retrieved seems to consist not of discrete elements of information, but instead of *relations* between the elements. Taking these various suggestions into account could also contribute towards a better understanding of long-term memory.

**References**


**The Nature of Precognition**


Puertomar, 11-5-D, Valdelagrana
11500 Puerto de Santa María
Cádiz, SPAIN
rjontaylor@hotmail.com

**Acknowledgements**

I would like to thank the Editor and the two referees for their valuable comments on the earlier drafts of this paper.

**Abstracts in Other Languages**

**German**

**Das Wesen der Präkognition**


Spanish

LA NATURALEZA DE LA PRECOGNICIÓN

RESUMEN: En este trabajo se describe una teoría que explica la precognición, literalmente el “pre-conocimiento” de la información contenida en el cerebro del percipiente en el futuro como una relación con su futura experiencia del evento. La teoría se basa en el modelo de universo bloque, en el que existen eventos pasados y futuros en el continuo espacio-tiempo, como es requerido por la teoría especial de la relatividad. La teoría de orden implícito de Bohm es compatible con un modelo de este tipo y sugiere que si estructuras similares se crean en diferentes lugares en el espacio y el tiempo, las estructuras resuenan con una tendencia a ser más similares entre sí. Los principios se aplican a los patrones espaciotemporales neuronales que se activan en el cerebro. Considero a la precognición como el fenómeno fundamental de ESP manifestado como la transferencia de información del cerebro en el futuro al mismo cerebro en el presente. El modelo permite también la posibilidad de contactos con otros cerebros, y estos contactos pueden producirse sincrónicamente o en momentos diferentes. Sin embargo, la teoría no admite en lo absoluto el contacto directo con objetos o acontecimientos externos. El mecanismo se aplica a los experimentos de precognición y explica las anomalías encontradas en los resultados.

French

LA NATURE DE LA PRECOGNITION

RESUME : Cet article décrit une théorie expliquant la précognition comme étant littéralement une « pré-cognition » de l’information contenue dans le cerveau du percipient dans le futur – un lien avec son expérience future de l’événement. Cette théorie est basée sur le modèle de l’univers-bloc, dans lequel les événements passés et futurs existent déjà dans un continuum espace-temps, ainsi que le requiert la théorie spéciale de la relativité. La théorie de Bohm sur l’ordre impliqué est compatible avec un tel modèle, et il suggère que si des structures similaires sont créées en différents lieux de l’espace et du temps, les structures ressonnent l’une avec l’autre et montrent une tendance à accroître leur similitude. Les principes sont appliqués aux patterns spatio-temporels neuronaux qui sont activés dans le cerveau. La précognition est interprétée comme le phénomène fondamental de la perception extra-sensorielle. Elle se manifeste comme un transfert d’information en provenance du cerveau du futur vers le même cerveau dans le présent. Le modèle permet également la possibilité de contacts avec d’autres cerveaux, et ces contacts se produiraient soit en temps réel, soit en temps différé. Toutefois, les contacts directs avec des objets ou événements extérieurs ne sont pas intégrés dans ce modèle. Le mécanisme proposé est étayé sur les expérimentations de la précognition, et il explique les anomalies apparentes découvertes dans les résultats.
REVEALING PSI SECRETS: SUCCESSFUL EXPERIMENTERS SEEM TO SUCEED BY USING THEIR OWN PSI

By Adrian Parker and Brian Millar*

ABSTRACT: A review of the literature makes a strong case for successful experimenters using their own psi to produce positive results either directly or in concert with that of their participants. The personal histories of well-known experimenters suggest that in some cases this capacity existed prior to their success as experimenters, and for others it seems to have occurred or reached awareness through experimentation. Examples are given of the successful performance of psi-conducive researchers when they were participants in experiments. Some hitherto unpublished aspects of the Gothenburg ganzfeld studies are presented that support these conclusions. Theoretical backgrounds in terms of current observational theory and current views of consciousness are briefly reviewed. Some ways of “forensically” or otherwise revealing the influence of experimenter psi are also presented. Experimenter psi is seen as having far-reaching implications for both psychology and parapsychology.

Keywords: psi, experimenter psi effect, replication, psychic experiences

This paper has its roots in a cooperative project which we carried out many years ago. At that time we initiated what has proved to be an all-too-rare cooperative enterprise between one of us (Adrian Parker) who was successful at finding high-scoring participants and therefore became a proponent in the field, and the other (Brian Millar) who consistently had the opposite experience and became a psi-skeptic. The results of this small project, which we will discuss more fully later, led to our being among the first to raise the issue of experimenter psi (Parker, 1978). Since that time the issue has steadily gathered momentum to such a degree that we now see experimenter psi as the driving force behind the replication and credibility problems that have come to bedevil contemporary parapsychology.

The situation is a serious one. If we are right and the pervasiveness of experimenter psi has become parapsychology’s best-kept secret, then it is what underlies the contemporary crisis of academic parapsychology. The crisis has led to a declaration of the field’s “demise” by the veteran and inexorable critic of the field, Ray Hyman (2012). Even one of the most renowned experts in our field, Donald West, recently raised some very legitimate concerns as to “why parapsychology has not become an accepted science” (West, 2012). In his reply to West, David Ellis (2012) mentioned the Gothenburg ganzfeld series as one of the three most “extraordinary findings” in the last 20 years. Perhaps it is precisely because of these extraordinary results that the series has given one of us some valuable experience and possible insights concerning experimenter psi effects, insights which are shared later in this paper.

Until now, relatively little attention has been given to the issue of experimenter psi. Although it was the subject of a comprehensive review by John Palmer (1997), his paper, despite being well founded, provoked little reaction. The reason may be that most psi-conducive experimenters preferred to believe in a social interaction theory of the experimenter effect rather than a psi-based one. Indeed, Palmer (1997) had already foreseen this when he remarked in his paper that amongst his academic peer group there would be much greater approval for possessing social skills than any proclaimed psychic ones.

Nevertheless, the time may now be appropriate for facing this issue head-on. Very recently, we discovered that Dick Bierman and James Spottiswoode independently of our input sought to challenge the psi-research community by claiming that the unresolved elusiveness of psi will hinder any future breakthrough. Bierman and Spottiswoode (2012) sum up their take on the situation as: “It seems to us that in most of our lab-research the participants are possibly just excuses for the psi-gifted experimenter to get rid of the responsibility for creating an anomaly” (p. 5).

Where Bierman and Spottiswoode as well as other authorities such as James Kennedy (2003) differ from us...
is that they see the evidence for this basic elusiveness in the paradoxical decrease in effect sizes with the increase in
the size of the samples used, and the supposed “meta-analysis destruction effect.” The alternative notion that we are
presenting is that this elusiveness can be understood as vacillations in the strength of psi effects deriving from the
influence of all the participants rather than in the very nature of psi. We, the joint authors of this paper, do, however,
differ in our opinions somewhat concerning the proportion of the results in parapsychology that can be attributed
solely to experimenter psi (emphasised by Millar) and the proportion that can be attributed to the participants (em-
phasised by Parker).

The lack of attention given in the research literature to the specific issue of experimenter psi stands in con-
trast to that given to the social skills of the researchers. It is often assumed this occurred because J. B. Rhine chose
to focus solely on the social interaction effects and to totally ignore the problem concerning the effect of the exper-
imenter’s own psi. This is, however, not entirely true. Rhine wrote about this issue in response to a direct question
from Parker when he, as a young student, wrote to Rhine, who replied to Parker’s query concerning experimenter psi:

I do not think the experimenter’s own psi ability has shown up to be as an essential factor, but I still think
we must assume that every experimenter has the ability and only needs to learn how he himself can liberate
and register it. However the very conditions that help the experimenter to liberate this might be the kind
that would help him to induce his subjects to perform successfully. (Personal communication, J. B. Rhine,
Sept. 5, 1975)

Rhine’s reply is revealing, and seen retrospectively, it can be considered well-founded. This is especially
the case in view of the little known fact that Rhine himself showed quite outstanding success as a subject-participant
in psychokinesis experiments (Averill & Rhine, 1945; Rhine, Humphrey, & Averill, 1945).

Lessons From History

Despite Rhine’s awareness of the role of experimenter psi, the initial research effort on experimenter effects
was to try to attribute success as an experimenter in parapsychology to the possession of interpersonal skills. As
early as 1937, the first two volumes of the Journal of Parapsychology contain three publications about experimenter
effects interpreted in this way. However, by the 1970s replication difficulties had become so pronounced that the
simple social interaction explanation of the experimenter effect was already being challenged. The crisis was, and
still is, epitomised by John Beloff’s now classic comment: “Rhine succeeded in giving parapsychology everything
it needed to become an accredited experimental science except the one essential: the know-how to produce positive
results when and where required” (Beloff, 1973, p. 291).

It was this statement, written in the 1970s, that motivated the efforts of Richard Broughton, Millar, and
Parker to attempt to come to grips with the experimenter effect, which was already seen as underlying the difficul-
ties in making progress. At that time Rhea White (1976a, 1976b, 1977) had published papers that brought together
the research on what we knew then about the experimenter effect. The evidence presented in these papers indicated
that the performance of participants was primarily influenced by the motivation and belief of the experimenter.
However, one of White’s papers suggested there was something more to it, namely the experimenter’s own psi abil-
ity and possibly even the checker’s psi-ability (White, 1976b). Like placebo effects in medicine, the experimental
ritual could be seen as merely a means of engaging the belief system of the designated “subject” in the experiment
to make things work. However in the case of experimenter psi, the role of the experimenter becomes a disguise to
avoid being the identified “psychic.”

Kennedy and Judith Taddonio were probably the first researchers to coin the term “psi-experimenter effect”
and explicitly focus on the experimenter’s own PK as the means of making positive results happen. They wrote:
“Parapsychologists should face the fact that an experimenter is typically more motivated than his subjects to achieve
successful results” (Kennedy & Taddonio, 1976, p. 3). As for the actual modus operandi of psi-mediated experi-
menter effects, these authors cast a spotlight on the experimenter’s unintentional PK as acting on the randomization
process. They believed the experimenter’s supposed PK is difficult to identify because the effect appears to be inde-
pendent of the complexity in the means of randomising the targets and decoys and is thus rather pervasive. Perhaps
the most striking evidence they present for this pervasiveness is the classic study by West and Fisk (1953) in which
the participants did not know of West’s involvement but nevertheless seemed to be reacting to his negative influence by scoring only on the cards prepared by Fisk and not those randomised by West.

It was also during this period that Parker (1976) introduced terms such as “psi-inhibitory” and “psi-conducive” to describe how the effect of the experimenter could be bi-directional, to inhibit as well as facilitate performance. At that time, Millar (1979) went so far as to suggest that the evidence for experimenter psi was so compelling that parapsychologists did not any longer need to recruit participants but could merely experiment on themselves. The suggestion was evidently not well received in the research community. The suggestion may have simply been too provocative, especially given that if “psychic experimenters” exist they may need to keep their identity secret and operate vicariously through their “subjects.”

For this reason it was thought that the use of a covert test might be the most suitable means of testing for experimenter psi (Parker, 1976). The covert test he used required marking a personal preference for certain words, some of which had been randomly selected as targets for each individual participant. The test was given to research parapsychologists attending a Parapsychological Association convention. Using pre-set criteria, three independent judges were able to identify 15 parapsychologists who they considered to be psi-conducive and another 12 they considered to be psi-inhibitory. It was found, as predicted, that the psi-conducive group scored significantly on the covert test of psi, whereas the psi-inhibitory group scored at chance level.

Back then, these results just seemed to be too good to be true, so they were never fully written up beyond a research brief and a chapter in a thesis. Parker was persuaded to see these successes as “beginner’s luck,” to use the words of John Beloff, his supervisor, or in the best case scenario, as being due to the investigator’s own psi, because back then Parker was running many experiments and succeeding in most of them. Parker produced at that time significant psi associated with hypnotic dreams, group expectancies concerning ESP, animal psi, the selection of high-scoring subjects, and his first ganzfeld series.

Attention at that time was still focused on finding the main basis for the experimenter effect in personality and interpersonal relationships. In this respect the classical study is still the one carried out by Charles Honorton and co-workers (Honorton, Ramsey, & Cabibbo, 1975). In this study, participants were randomly assigned to two groups, both of which carried out a precognitive task with a random number generator. One of the groups was treated in a friendly and supportive manner, whereas the other group was handled in an abrupt and unfriendly manner. The results showed the expected differences, with the friendly-treated group giving positive results and the unfriendly-treated group obtaining negative ones with the difference being significant.

This study has received relatively little attention compared to, say, the more contemporary experiments carried out by Richard Wiseman and Marilyn Schlitz (1998, 1999). In contrast to Honorton’s consistent success, Wiseman has a much acclaimed lack of success at psi testing. Nevertheless, Wiseman and Schlitz essentially replicated the study of Honorton et al. (1975) by finding an initially significant difference between Schlitz and Wiseman and relating this to their different (warm versus abrupt) styles of handling participants. In the oral presentation of a convention report on the first studies (Wiseman & Schlitz, 1998), Wiseman commented: “Either Marilyn [Schlitz] is cheating or something weird is happening!” This statement is not entirely lacking in irony given Wiseman’s reputation at the time as a security expert, in view of which it might be expected that the first cheating alternative would have been effectively ruled out. Ten years later, when taking into account the failure of a third and final attempt at replication, Wiseman regarded the series as a whole as not “stunning evidence” for psi (Ritchie, Wiseman, & French, 2012, p. 348). However, given the replication difficulties inherent in cognitive and social psychology (Editorial, 2013), the results might be considered “stunning evidence” for the difficulties in replicating exactly the same interpersonal conditions on three consecutive occasions. If we prefer the statement that “something weird” was going on, the crucial question becomes, was the weirdness due to the different styles of the experimenters or to some form of experimenter psi?

A similar impasse concerned the Watt and Ramakers (2003) study, which replicated the original 1975 study of Parker showing that the experimenter’s belief in psi influenced the results. When the authors attempted to look more closely at how this effect occurred, they found no personality or cognitive correlates that might hint at an explanation for the experimenter effect. The authors were left unable to decide between whether the effect was caused by experimenter psi, the believing experimenter eliciting psi from participants, or an interaction between experimenter psi and participant psi.

There is, however, one study that does address the psychology of the psi-conducive experimenter: the questionnaire study of professional psi researchers carried out by Matthew Smith (2003). Smith found that those
experimenters who considered themselves successful at testing for psi reported more often having their own psychic experiences, which led Smith to conclude that this gave them greater confidence in the success of their experiments.

Smith and Savva (2008) experimentally manipulated the belief in psi of experimenters who carried out an auto-ganzfeld study. However, the significant difference in scoring rate between experimenters did not relate to the manipulated expectancies. Rather surprisingly, it would appear that no reported attempt was made to assess whether or not the manipulations of belief went against or reinforced the experimenters’ prior beliefs, which of course could have affected the outcome. More to the point, there was no attempt to see if the successful experimenters could be distinguished by their own performances in the psi tasks. As this was for Smith and Savva a rare successful experiment, we might have expected a follow-up, but both researchers left the field disillusioned. Smith, apparently disillusioned with academic life and enthused by his performance in the TV series Most Haunted, left the university with an aim, according to his website, to become his own media psychic (Smith, 2011). More bizarrely, his co-experimenter, Louie Savva, also left the field but opened a website on which he declared “everything is pointless” in life and “parapsychology is a fake science” (Savva, 2006).

Perhaps it is because of this reluctance to deal with the messy implications of experimenter psi that the focus on psi-mediated experimenter effects has not been sustained. Today it appears to be true that most academic parapsychologists, at least in the UK, look at psi as being normally distributed amongst the population and to be studied like any other psychological ability. From this point of view, participants are regarded as having measurable abilities, which can be studied objectively by a detached observer and which, like other abilities can, with appropriate motivation, be engaged in the task at hand. This view tends to dominate, perhaps because it is seen as a ticket to academic respectability in UK psychology departments. But questions can be raised as to whether this approach has really worked and, more to the point, whether it is supported by evidence.

The research evidence for this, in our opinion a rather presumptive viewpoint, is far from convincing. Kennedy (2000) addressed this issue in his paper, “Do People Guide Psi or Does Psi Guide People.” He reviewed the evidence for strong genetic loadings for traits, such as absorption, for which there is evidence of a positive relationship with psi ability. Kennedy (2000) concluded: “These results provide a basis for the hypothesis that psi-ability (whether as experimenter or subject) has a substantial innate or genetic component” (p. 136).

Naturally, the question pertinent to the above-mentioned research strategy is how common the supposed ability is. Kennedy estimates that only between 10 and 15% of the population have genuine psi experiences. If we take into account the figures from the survey of psychic experiences carried out in the late 1970s by Palmer (1979) and select those who have frequent psychic experiences (5 or more), which might be considered a selection criterion for succeeding in psi experiments, 17–19% of the survey responders would fall into this category. If we take Kennedy’s estimate that only 10–15% of these have genuine psi (although this may be different for multiple experiences), clearly we are left with a small percentage, probably no more than 3%, who are suitable for research.

If multiple genuine psi experiences are rare, this would explain the unreliability of findings in parapsychology based on group testing or on unselected participants. In this vein it should be mentioned that the success at Gothenburg with the ganzfeld was partially attributed to the careful selection of individuals reporting frequent psychic experiences (Parker, 2000). It is also the rationale behind our current twin studies. Even amongst identical twins, we would estimate that less than 10% experience apparently genuine psi at the level of potentially being able to reproduce it under laboratory conditions (Brusewitz, Cherkas, Harris, & Parker, 2013; Jensen & Parker, 2012; Parker & Jensen, 2013).

This naturally leads us to wonder if psi-conducive experimenters belong to a self-selected group. If so, that raises two questions: Do the successful parapsychologists have their own psychic experiences and, if so, do these precede their success as experimenters? If the latter is the case, it would give some support to the idea that successful parapsychologists may be using their own psychic ability to produce the significant results.

**Personal Psi Experiences Amongst Researchers**

It seems natural to begin the quest for answers by asking, what is the evidence for psi experiences occurring in the personal lives of well-known parapsychologists? The task is made difficult because it is usually considered to detract from objectivity for psi researchers to admit to personal experiences.

An exception is Haakon Forwald, perhaps the most successful PK experimenter of modern times, who in the 1960s reported a long series of successful PK experiments in which he was his own subject. He gave Robert
McConnell an example of the kind of psychological strategy he constantly improvised: “Just in advance of each release (an assistant) would playfully push [a toy plastic motor car] on the apparatus table in the direction in which they desired the cubes to go. This seemed to work well for a few releases before the effect disappeared” (McConnell & Forwald, 1967, p. 206). Forwald’s psychic ability may have been more general than influencing dice and cubes. According to the memory of Forwald’s granddaughter, whom Parker interviewed, Forwald had a reputation for diverse psychic effects, for instance in the house he built, which was later reputed to have become haunted.

Another Scandinavian, Martin Johnson, is a striking example of the combination of academic success as a psi experimenter (Johnson held for 13 years the Western world’s only state funded professorship in parapsychology) while apparently possessing some form of psychic ability. The claim has some documentation. Johnson was raised in Northern Sweden in the Sami culture, and he recorded many examples of his own premonitory dreams, of which at least one must be considered well-documented: the shooting down by the Russians of the DC-3 spy plane (see Parker & Möreck, 2011).

Johnson was successful in a wide range of psi-testing situations, including the use of the pioneering Defense Mechanism Test (DMT) to predict ESP scores. The DMT was once highly regarded as a strong candidate for the repeatable experiment in parapsychology, but its success was entirely dependent on Johnson’s involvement. In an attempt to break this deadlock, Parker arranged a supervised session with Johnson, who jokingly introduced the session with the words “And now we shall have a séance!” In this context it is interesting to recall that a later reanalysis indicated the hits correlated significantly with the DMT scores most often when Martin Johnson deviated unaccountably and markedly from his set pattern, which followed the scoring rules of the DMT manual (Haraldsson, Houtkooper, Schneider, & Bäckström, 2002). The scientist can be seen as a modern form of shaman with set rituals to make the phenomena happen; perhaps the use of the DMT fulfilled this need in Johnson (Puhle & Parker, 2003).

There are several further outstanding examples of leading researchers in parapsychology who alluded to or documented their own psychic experiences. In his article “Are We Shamans or Scientists,” Rex Stanford (1981) writes: “Since I myself have been successful as a subject in several ESP studies and have had many spontaneous experiences which seemed psi-mediated, I feel I am in a position to say that these experiences have had a definite value to me in planning my experiments. Moreover in trying to create a psi-favorable setting for a study, I never ask my subjects to undergo conditions which I have not experienced myself and judged to be favorable in this regard” (p. 62).

This latter principle is one which Parker has used in his own experiments and explains why some of the best of the hits he cited (Parker, et al, 2000) to illustrate the success of his “real time digital technique” (elaborated on later) occurred when the experimenter took on the role of participant.

In explicitly answering the question he poses, Stanford divulges the best-kept secret that we are referring to. He writes: “Indeed every experimental parapsychologist personally known to me who has a record of obtaining significant results fairly regularly—results which have à priori credibility and are obviously not the results of post-hoc analysis—also has a personal record of having been a quite successful subject in psi studies” (Stanford, 1981, p. 62).

A striking example of what Stanford is referring to involves Honorton, who arguably is the most successful psi experimenter of recent times. Whereas nothing seems to have been written about him having subjective psi experiences, he was undoubtedly successful as a participant in PK studies (Honorton, 1976). Following Honorton’s premature death, Stanford, in his tribute to Honorton, noted how the success of Honorton’s ganzfeld work could have depended on unconscious experimenter psi in choosing targets selected by a random event generator. Honorton had a previous track record as a subject in successfully influencing random event generators. To reduce such an influence, Stanford had urged Honorton to use a more complex means of randomization that would employ computerized sampling of random number tables, as this might be less amenable to psi influence. Apparently this had been planned but was never realised because of Honorton’s premature death (Stanford, 1993).

Like Honorton, William Braud, one of the co-founders of the psi ganzfeld, achieved an extraordinarily consistent level of success. In an article entitled “Honoring Our Natural Experiences,” Braud reveals a series of his own subjective psi experiences. Some of these were of an apparent telepathic nature, whereas others related to synchronicities and PK (Braud, 1994).

Unfortunately, there is a lack of documentation of the personal psi experiences of the other successful psi researchers, who are relatively few in number (e.g., Helmut Schmidt, Kathy Dalton, Dean Radin, Rupert Sheldrake). We do know, for instance, that Schmidt was successful as his own subject in random number generator experiments (see Kennedy, 2000, p. 134), but little is known about his personal experiences. Radin has communicated to us that
his personal experiences occurred only after becoming a successful experimenter (Personal communication, D. Radin, March, 10, 2014), whereas Schlitz (2009) considers her ability to be the result of both her personal experiences and of her success as an experimenter.

This raises the question: Could successes arise as an outcome of laboratory work rather than as a precursor of it? This viewpoint seems supported by the lives of J. B. Rhine and Gaither Pratt. Although, as mentioned earlier, Rhine succeeded as a PK subject, there is nothing in their respective biographies (Berger, 1988) to suggest that either of them possessed any psychic ability, and this is confirmed by a statement to Parker by Rhine’s daughter Sally Rhine Feather (Personal communication, S. R. Feather, May 20, 2013) and Jurgen Keil (1979), who both knew much about Gaither Pratt’s life history. Of course, it is still plausible that there existed a latent predisposition for psi that required the laboratory context for its expression and potential recognition. What Kennedy (2000) has called “ownership resistance” in relating personal experiences could certainly have hindered the full recognition of experimenter psi.

There are examples of other researchers who reported seemingly psychic experiences that inspired their interest but which did not seem to profit them in terms of their profiles as psi experimenters. Kennedy (2000) reported having psychic experiences during a limited period in his life, White had a near-death experience (2014), and Susan Blackmore (1987) had an OBE. However, the transitory and rather specific nature of aspects of these experiences sets them apart from the frequent and more integrated experiences that appear to be most typical of psi-conducive experimenters.

It is clear that the complex question of the direction of causality of psi (i.e., which comes first, the experience or success in the laboratory) is not easily resolvable as simple one-way causality. Nonetheless, the answer has implications for the replication problem in parapsychology.

**Experimenter Psi as a Source of the Contemporary Replication Problem**

There may exist further living examples of psi-conducive experimenters. We are thinking of the consistent successes of Roger Nelson and Spottiswoode in confirming their hypotheses. It is beyond the scope of this article to go into the debate on the findings relating to the “Global Consciousness Project,” but suffice it to say that Edwin May and associates have argued that the findings are due to an experimenter psi effect mediated by Nelson. The interested reader is referred to the contributions by May and Spottiswoode (2011a, 2011b), Nelson (2011), and Bancel (2011). Given this, it is perhaps ironic that May and Spottiswoode might themselves be psi-conducive experimenters. During the period when parapsychologists’ research on sidereal time was much in vogue, May was given access to data from the Gothenburg ganzfeld studies, and the analysis apparently gave significant results that further supported the sidereal hypothesis (Spottiswoode & May 1997). However, given that there are so many potential variables that could influence psi scores, it seems intuitively highly unlikely that the earth’s rotation relative to the stars could dramatically affect such scores, especially taking into account that in the case of the ganzfeld, the psi effect constitutes only about 10% of the score variance. It is also worth noting that this hypothesis was formulated when Spottiswoode found a sidereal time relationship with psi scores. It was a relationship that remarkably would turn out to have predictive value and came to dominate research for nearly 10 years.

Another area which appeared promising during the 1990s involved the Princeton Engineering Anomalies Research (PEAR) group, which claimed that lengthy testing showed that “operator intentions” could have small influences on the outcome of random generators (Jahn, Dunne, Nelson, Dobyns, & Bradish, 1997). Although this was initially meant to be a replication of the earlier Schmidt work with random number generators, the highly significant patterns that emerged had, in contrast to the Schmidt findings, extremely low effect sizes. Such meagre effects never impressed critics who would merely argue “error some place.” One of them, James Alcock (1988), pointed out that although many participant-operators were involved, the results appeared to be dependent on one of the main experimenters when she was in the role of the key operator, namely Jahn’s co-worker and partner Brenda Dunne. Apparently, the PEAR group has refuted this, showing that the results still reached significance when the scores with Dunne as the lead experimenter were removed (Dunn et al 1988). Yet, as far as we are aware, amongst the numerous analyses, none is reported for experiments in which Dunne was not involved in the team of experimenters, as the Princeton studies were always conducted as a team effort. Whatever is the case, the PEAR work did not replicate (Jahn et al., 2000).

A contemporary line of research concerns the extraordinary successes of Rupert Sheldrake with experiments
on the sense of being stared at (Freeman, 2005). In contrast to the sidereal time hypothesis, which is entirely empirically derived, the hypothesis concerning the sense of being stared at can be said to have ecological validity in that it is a rather common belief (Sheldrake, 2005). Nevertheless, Sheldrake’s impressive series of positive findings may have replication difficulties, as we will now detail.

There is a debate over how much of the success in the staring studies can be explained by a combination of response bias and feedback (Lobach & Bierman, 2004). Experiments reported by Marks and Colwell eliminated the feedback and still obtained positive results. However, in their follow-up study, along with adding improvements, Marks and Colwell substituted a sender with negative expectancies about the success of the experiment. The latter change led to irresolvable differences of opinion as to why this experiment failed to replicate their former success (Marks & Colwell, 2000). In a further study, Stephan Schmidt and co-workers also failed to replicate the Sheldrake findings, but in their case no information was provided about the expectancies of the two “trained starers” they used (Müller, Schmidt, & Walach, 2009). Therefore, at present, a conservative conclusion is that the positive results appear to depend on the involvement of Sheldrake or at least of a psi-conducive experimenter.

The final apparent example we cite can be found amongst our Stockholm colleagues, who carried out a series of studies with group testing of ESP that have continued for about 10 years. Although the researchers have consistently failed to find the predicted effects, they have also consistently found large, sometimes even astronomical, post hoc effects. Clearly, post hoc effects lack validity when, as in this case, the follow-up fails to confirm the effects that had previously been found. However the peculiar aspect in this case is that large post hoc effects kept appearing, thereby encouraging the now 10-year chase after psi to continue (Dalkvist, W. Montgomery, H. Montgomery, & Westerlund, 2010). The critic would be tempted to ask: Are they chasing post hoc effects in the form of chance-determined group differences or are they chasing the influence of experimenter psi on group performance? In deciding between these alternatives, it is worth noting that the main hands-on experimenter in the series, Joakim Westerlund, admits to having a propensity for synchronicities.

We stop here, as Millar (2012) cautions us not to give too much attention to the investigation of current cases of apparent experimenter psi, lest this exposure of effects interfere with its continued manifestation. Millar sees the effects as pervasive and psychologically elusive.

Parker differs somewhat here from Millar (see Millar, 2012; Parker, 2013a). He emphasizes that it is counter to the evidence to conclude that all or even the majority of the successful experiments in parapsychology are entirely due to experimenter psi in some form or other. It seems rather self-evident that if there are psychic experimenters then there have to also be psychic participants. In many cases, it may be difficult to discern if there is a single psi source, as it may be a conjoint effort. This appears to be the case when Parker, after testing about 25 individuals, found 2 who could apparently demonstrate the searched-for psi ability. One of them continued to score high with Parker as the experimenter. However, when Millar took over as experimenter, he scored significantly with Millar but no longer with Parker, although it should be added that the difference in scoring between the two experimenters was small and nonsignificant. The remarkable aspect is that with the exception of some very short-lived personal successes at card guessing during his youth, Millar has during many years of effort never before this experiment or since succeeded in getting significant results with psi testing of participants.

The second of the participants found by Parker who maintained her high scoring during a long period produced a very high score (15 hits of 25 with an MCE of 5), with all the remaining 10 calls being displacements, in the presence of John Beloff as an observer, despite his reputation of being a psi-inhibitory experimenter (Parker, 1975). These scenarios seem to fit with what Rhine implied in the correspondence quoted earlier, namely that the experimenter in this case somehow enabled potentially high-scoring subjects to express their own ability. It is also noteworthy that the personal history of the two high scorers suggests that they both had spontaneous psychic experiences prior to their laboratory performance and these were congruent with this performance (Parker, 1975).

As regards Parker’s personal history of subjective psi experiences, it is perhaps worth noting that his mother had a reputation for being “psychic” (although he was too young at the time to recall this). While Parker has experienced several apparently psychic events, most of these occurred after his research career began. Working in an academic setting has meant developing skills in seeking normal explanations and seeing research as a means of reconciling his skeptical and open-minded sides. Although he attempts to be open to such experiences, and at one level accepts them, he prefers to give more credibility to those relatively rare occasions in which he had been a participant in experiments where normal explanations could be excluded. Some of these occasions are mentioned below. His personal experiences confirm those of the laboratory insofar as they seem to occur only when the ambience favors...
them. It is for this reason that he largely restricts his own participation to the occasions when he feels confident of succeeding, such as testing new experimental designs he is enthusiastic about.

Millar’s thesis is that experimenter psi is the major interactive and rather pervasive factor that explains many of the inconsistencies and much of the elusiveness of psi. We thus are not entirely in agreement. Some of the data presented below by Parker, although largely anecdotal, suggests to him that the situation may be more complex.

**Experimenter Psi Effects in the Gothenburg Ganzfeld Series**

The qualitative analysis of the “Gothenburg best ganzfeld hits” (Parker, Persson, & Haller, 2000) contains some 20 qualitatively real-time correspondences between the target clips and the mentation reports that were deemed to be “best hits.” Although both Honorton and Dalton had reported such qualitative correspondences, the unique feature of the Gothenburg ganzfeld was the recording of these hits in real time with events as they were being shown in the film. This means the receiver’s mentation was superimposed on the possible film targets, but only the actual target had the real-time effect. What is relevant in the present context is that half of the best hits presented in the published report were actually derived from those trials involving either Parker or the two other researchers as receivers. In assessing the value of this statement, it is essential to bear in mind that the target pool had been arranged prior to the experiment by an assistant. This means that the experimenters in the role of participant contributed their sessions without having any prior knowledge of the target pool of film clips.

It should also be emphasized that because the contribution of the experimenters was only a few sessions, their participation made no essential difference to the overall significance of the quantitative results.

One of these experimenter-participants was Dalton, who had run the very successful Edinburgh ganzfeld series. Another experimenter-participant, Annhild Haller, who contributed to the publication on impressive real-time hits contained in this report, was a professional psychologist who was a highly successful receiver in the earlier series of ganzfeld experimentation. It was then the custom at Gothenburg, when running experiments with new designs, that the technique would be tried first by the experimenter in the role of receiver. Parker had contributed four trials within a single session that belonged to a special pilot series in which four targets were used in random order across sessions. The outcome of this one session (the only one Parker contributed to the series) was that all four of the transcripts from Parker’s mentation enabled correct order placements to be made by him, and all appeared to show striking real-time correspondences with the target clips (Parker & Westerlund, 1998). Three of these real-time hits were included as illustrative examples in the summary publication about qualitative hits (Parker, Persson, & Haller, 2000).

Once again it is important to emphasise that no claim is being made that the Gothenburg ganzfeld results were entirely due to experimenter psi. Only three trials in the major ganzfeld series of 150 trials were contributed by two of the experimenters. However, about half the selected best hits were from experimenters performing as participants. Three of the remaining “ordinary” participants, it should be added, returned to the lab for a second, or in one case a third, session to repeat their initial successes. What we claim is that experimenter psi probably played a meaningful and important role in enabling the overall success of the experiment.

**Theoretical Issues**

If we are going to penetrate the issue of experimenter effects at a deeper level, we need to discuss the basic theoretical issue about the nature of psi: Are we dealing with purely quantum-level statistical effects, the type of observer feedback that forms a basis for observational theory, or something more, requiring that psychological factors are a decisive influence.

The qualitative analysis of the ganzfeld work described above, using the recording of hits in real-time, suggests at first glance that we are not merely dealing with statistical anomalies or simple observational effects. This is because it is difficult to conceive how such effects would enable participants to give real-time hits while they appear to be describing the external events displayed on the computer screen, sometimes sequence by sequence, just as they happened. One of the best indications of a causal link was when Parker, without divulging this plan to anyone, replaced, in the middle of a session, the designated sender with a close friend of the receiver. The idea was to see if this “intervention” would influence the mentation report of the receiver. At the exact time this substitution was made, the receiver remarked: “Where have you been?” She later reported that these words had spontaneously
entered her head without any imagery. Naturally, a counterargument to that of psi would be that this is merely sub-
jectively validating what is sought. What speaks against this argument is the real-time aspect and that nothing else
during the 30-min session appeared to relate to the sudden entry of her friend. Moreover, a further incident occurred
with this participant when the recording tape in the room suddenly stopped; this occurred a few seconds after she
said spontaneously during the mentation period, “Change the tape!”

This was an exciting and inspiring time for us because it appeared that the “real-time digital ganzfeld” was
given us the much-sought-after portal through which we could observe some of the psychological processes that
mediate real-life psi functioning. What we observed suggests that so-called top-down processes were involved in
interpreting psi-derived images. (Top-down processes are the processes involved in forming our normal perceptual
images of what we call reality. They occur by interpreting the basic information from the senses, using the uncon-
scious influence of memory images and expectancies.)

Regrettably, the ambience necessary for the good teamwork that seemed to be enabling the success did not last. The attitude towards university-based parapsychology in Sweden unexpectedly changed for the worse and became hostile. Tensions grew within the research team, culminating in a crisis over whether or not the primary analysis should be based on receivers’ own evaluations of the ganzfeld mentation (advised by Parker) or on that of an external judge (preferred by another experimenter). Finally, a reluctant agreement was reached prior to running the experiment to use both sources of assessment (but for the purpose of the doctoral thesis, the overall results were primary). The outcome was that the judge’s scores were at chance level and psi-missing occurred in the form of a statistically significant negative score for the receiver evaluations (Goulding, Westerlund, Parker, & Wackermann, 2004).

The above experience may be indicative of how the ambience and the whole situation influence psi perfor-
mance. Four years later, when the university prospects for parapsychology in Sweden had drastically worsened, we
attempted to repeat the earlier ganzfeld results with a much more complicated procedure that enabled potential tar-
get film clips to be primed before being selected. We obtained only chance scores (Parker & Sjödén, 2010). Perhaps
symptomatic of the prevailing pessimism, Parker did not first try the procedure out using himself as participant.

Whatever their nature, none of the earlier type of remarkable real-time correspondences reappeared in these
new data. If they were mere subjective validations, as my psi-critic colleague Westerlund would have expected, then
they should have. It was Westerlund who carried out the “remarkable” reassignment of the mentation reports from the
above psi-missing experiment, so that they were now matched to new randomized targets. The fact that the new
dummy matches did not differ in quality from those of the original hits indicated to Westerlund that the real-time
hits were all mere subjective validations. There was, however, a flaw in this argument, namely that these particular
original hits were few in number and were from a series that had not given psi-hitting, but psi-missing (See com-
ments by Parker in Westerlund, Parker, Dalkvist, & Hadlaczky, 2006).

The story of the Gothenburg experiments reinforces the idea of the “elusiveness of psi” or even “the trick-
ster effect” of psi. The expression “morphing” is suggested here as an alternative term that emphasises how psi does
not necessarily evade or hide but rather changes its form. The use of the term “morphing” avoids the anthropomor-
phising of the trickster but still describes how psi is neither robust nor an ability that can manifest in the same form
irrespective of the conditions. Instead, it would seem that psi arises as an expression of the conscious, and perhaps
more so, the unconscious mind influencing random data.

Meta-Analysis and the Psi-Based Experimenter Effect

A recent review by Kennedy (2013) contains some challenging comments relevant to experimenter psi
effects. In asking the question, “Can we move beyond the controversies of retrospective meta analyses,” Kennedy
argues that meta-analysis has failed to resolve the issue of the existence of psi. He argues that this is because the
predominance of small samples in psi research easily loads the analysis with biased data and thereby magnifies the
influence of experimenter psi and other extraneous variables while at the same time disguising it. The critic can
evertheless seek refuge in such variables by renaming ESP as “error some place.” For this reason, Kennedy argues
strongly against the use of small samples and advocates a return to large-N, carefully controlled experiments.
Kennedy’s critique ought to in principle create a rapprochement between parapsychologists and critics such as
Hyman. To the dismay of many parapsychologists, Hyman suddenly disavowed the use of meta-analysis as a means
of resolving the issue of psi’s existence. In making such a move, he seemed to no longer be just moving the goal
posts, but turning them around. But seen from Kennedy’s perspective, there is justification for this sudden move.

One important aspect of parapsychological meta-analyses is the enigmatic lack of a relationship between sample size and statistical significance in terms of the $z$ scores. Kennedy calculated the actual power of the various meta-analyses to reliably demonstrate this finding. Using simulations, he was able to show that of the four meta-analyses carried out on ganzfeld data, only the first by Honorton (1985) and the last by Storm, Tressoldi, and DiRisio (2010) had the power to reliably evaluate the correlation between sample size and standard deviation. Honorton’s meta-analysis failed to do so, but this could have been due to the small sample sizes of the studies, which rendered them sensitive to extraneous variables. The confirmation of a relationship found in the latest meta-analysis gives us new confidence that the findings concerning the ganzfeld as a psi-conducive state actually do follow the laws expected for “normal phenomena.” Indeed, given the ecological validity of the ganzfeld in reliably reproducing strong results in the laboratory, and what we believe we know about psi-conducive states, it would be surprising if this expectation is not fulfilled.

Kennedy applied his simulations to a range of findings in contemporary parapsychology. One of them could have worrying implications for psi research. For PK research there is the absence of a relationship between sample size and $z$ scores. Also, the PK effect is apparently unrelated to the complexity of the target goal. For Kennedy, this means that goal-oriented psi-experimenter effects could occur independently of sample size and target complexity. PK by an experimenter could thus influence the participants’ responses as well as the random events in an experiment.

This led Kennedy (2013) to conclude: “The ganzfeld procedure could make a participant more susceptible to psi influence by the experimenter rather than facilitate psi by the participant” (p. 14). But is Kennedy’s argument entirely consistent?

The large-scale meta-analysis of Storm et al. (2010) indicates that the results for the ganzfeld as a psi-conducive state, are, as they should be, dependent on sample size. Moreover, they are relatively independent of the experimenter. It is of course justifiable in principle to follow Kennedy’s advice of using larger sample sizes in evidential studies to give the appropriate power. Yet from a pragmatic point of view, given the difficulty in funding large-scale parapsychological studies, and the above-mentioned necessity of selecting psi-gifted individuals, even small-scale studies, as long as extraneous variables are controlled for, may not be as inappropriate as Kennedy concludes. Ultimately, the crucial need is for large effect sizes, which can teach us something fundamentally new about the phenomena.

Explaining the Experimenter Psi Effect

Palmer devotes a major part of his earlier-mentioned review of experimenter psi to discussing its mechanism. One of the principle points he expresses is: “The experimenter does exactly what he or she asks the subject to do, except the experimenter does it non-intentionally and unconsciously” (Palmer, 1997, p. 117). As he notes, it is the intensity of the experimenter’s psychological involvement and intentions that seem to be crucially important. If this is so, it would argue against the application of any simple observational theory to explain the results, as intentions are not always conscious, because we generally lack full insight into our motives. Modern cognitive psychology, although distancing itself from the theories of Freud, accepts that we can have unconscious intentions and even dissociated intentional states (Hassin, Uleman, & Bargh, 2005).

Palmer also lists motivational factors, such as confidence of success in the psi task, as being crucial to the experimenter psi effect. Indeed, psi may function to some extent in the same way as “self-efficacy” in sports psychology, that is, as a skill which is based on the experience of previous successful performances. The form of expression that this “psi-efficacy” takes may thus determine the different outcomes of the experiments of psi-conducive and psi-inhibitory researchers. Like self-efficacy in high-performance sports, the psi-efficacy of the experimenter may be dependent on previous experience of successes and on the ambience, together with the supportive enthusiasm, of the testing situation. This conclusion seems surprisingly consistent with J. B. Rhine’s statement (given earlier) in which he said: “The very conditions that help the experimenter to liberate this his own psi might be the kind that would help him to induce his subjects to perform successfully.” A positive zeitgeist and the zeal of the researchers appear to be major conditions, and they may in part explain the astounding early successes of Rhine and Pratt.

Although the classical form of observational theory seems inadequate to explain the psi-based experimenter effect, the minimalistic form of observational theory advocated by Millar (2012) may be more successful.
“Minimalistic” refers to the idea that the farther the time of observation is from the initial experiment, the more the effect of the observation on the scores will decrease. For example, a hands-on experimenter’s influence is going to be greater than that of a reader of the report, especially when the final results are read as part of a meta-analysis.

It is here that the present authors have different opinions as to how the experimenter effect may work. Millar places emphasis on the contemporary concept of “decoherence,” where each event is apparently part of a larger influence in the transition towards macro events. For Millar it is the time of feedback, rather than the observation itself, that is the crucial influence.

There are, naturally, other views. For instance, Stuart Hameroff sees quantum decoherence as giving the sudden “pling” of reflective awareness via the activity of quantum processes in the microtubules of the neurons of the brain (Hameroff & Penrose, 1996). Parker points out that psi cannot be lifted out and separated from consciousness as an ability. He supports the view that although psi seems to be unconscious, this is so only in our normal waking state. Other states, such as the ganzfeld, hypnosis, meditation, and psychedelic drug states, have consistently been documented as enabling access to a wider spectrum of consciousness, including a psi mode of functioning (Parker, 2004, 2010b).

For Parker, observational theory cannot alone account for laboratory successes. Whereas Millar views observational theory as a sequence of non-causal correlations, Parker notes that some versions of observational theory propose that psi capitalizes on temporary deviations from randomness in the generation of target sequences (Houtkooper, 2002). According to this version, a form of psi-based intuitive scanning identifies the biased sequences, which yields correct responses to the targets in such sequences. Instead of intuitive scanning, other versions postulate that PK from the participants or the experimenter causes the sequences to conform with the target guesses. Whatever the version, if the successful target sequences in ESP experiments are analysed, they should show hitting during the periods with nonrandom sequences. The problem is that there is apparently no evidence for this.

One of the few recent attempts to study the nature of the experimenter effect was by Chris Roe and his co-workers at Northampton and Edinburgh. This approach, which we might identify as the “British School,” emphasizes the social interaction theory of experimenter effects. This means in practice that measures of the experimenter’s social skills should predict participants’ psi scores. However, in this case, despite numerous measures of rapport and quality of interaction, the researchers have been unable to find any significant differences in the experimenters relating to success in experimentation with the ganzfeld, thereby confirming the findings made in an earlier decade by Parker, Millar, and Beloff (1977) also using ganzfeld methodology. The conclusion Roe and his colleagues draw from their own research strengthens the main thrust of our argument. They write: “Perhaps extremely highly motivated experimenters with extremely high expectations of success who do not question the existence of psi at all, and who have readily integrated it into their daily lives, are required” (Sherwood, Roe, Holt, & Wilson, 2005, p. 169).

Contrasting with the “British School” there is another approach, which might for convenience be labelled as the “Continental School,” because it includes Dick Bierman and Walter von Lucadou. According to this school, nature operates on the basic limiting principle of prohibiting a level of replication or a degree of correlation that would enable psi to produce a causal signal within the system. This is because strong causal effects ultimately enable the control of psi, and the final result of this cause would threaten the integrity of the system.

Contrary to this theory, there are examples where a complete signal has successfully been transferred or the information influencing present-day decisions has apparently been retrieved from the future. A well-known example of the former is James Carpenter’s (1991) success in using a series of majority vote outcomes in a forced-choice ESP experiment to identify the word “peace” with complete accuracy. Keith Harary and Russell Targ successfully predicted the outcome of the silver market by remote viewing hidden objects that were to be associated with future ups and downs in the market. Although one attempt to replicate this failed, two other forms of replication apparently succeeded (Targ, Katra, Brown, & Wiegand, 1995). Moreover, the recent paper by Kolodziejczyk (2013) reports the results of a lengthy experiment using associative remote viewing to predict the future outcome of trading. By this method 60% of the actual trades were profitable, giving a net profit of $146,587.30.

There are other problems with the observational theories. In the opinion of Parker, the exclusively quantum theories of psi fail because they do not take into account the rich nature of spontaneous experiences that formed the work of the pioneers of psychical research. Neither do they relate the phenomena to normal psychological processes.

In this latter respect, the work of Edward Kelly and Emily Kelly (2007) is to be applauded because of their enormous achievement of integrating the insights of Myers and James on psi, dissociative states, hypnosis, and subliminal perception with the contemporary conceptions of the cognitive and the emotional unconscious. This work
The mind is described by Carpenter (2004) as “a zone of conscious awareness with a subliminal surround that precedes and conditions consciousness in many complex but systematic ways” (p. 19).

This view led Carpenter to his hypothesis of functional equivalence, meaning that the mind is pragmatic about drawing its information from various sources. This means that when sensory information is insufficient for decision making or even normal perception, the mind uses extrasensorially derived information, along with subliminal information and implicit memory, to fill in the gaps. Accordingly, in a guessing situation, choices are never random but are steered by all the competing processes that in principle are open to psi. However, attention also has a decisive role, and psi-missing can occur when there is avoidance of dealing with the uncertainties of the psychic information. Under these circumstances, what may be psi-based is set aside by choosing what may seem to be the safe bet. In such instances, the net result will be, not a chance result, but a negative score or strange reversals, as in the enigmatic Woodruff and Dale (1950) experiment.

Yet this model, even with its hint of panpsychism, fails when seen from the opposite direction: There is no bridge to modern physics. It can be speculated that such a bridge may be found in nonlocal effects if their presence in the brain is confirmed. Until this occurs, it must be recognised that at present psychology still operates on rather outdated models of physics and neurophysiology that exclude psi. This does not bode well for the integration of parapsychology with neuroscience (Parker, 2013a). Perhaps a third alternative for ESP needs to be promoted that might lay the foundation for a change of attitude toward the field. Rather than defining ESP provocatively as “extrasensory perception” or cynically as “error some place” and then trying to hide psi as “anomalous phenomena,” in view of the above considerations it would be better to use a more honest and appropriate term, namely, “extraordinarily synchronous phenomena.”

Handling the Experimenter Effect

Although he and I do not entirely agree about the pervasiveness of the experimenter psi effect, Millar (2012) suggests ways to specify when it occurs. He favors what he calls “forensic means,” by which the identity of the “psychic” or psi-source in a given experiment can later be revealed. For instance, Millar predicts that the psychic experimenter will cause mirror-image effects in the control groups, depressing their scores as well as elevating those of the experimental groups, in order to reach statistical significance. This would mean that a possible sign of experimenter psi would be the absence of significance for the entire sample but the presence of a significant difference between groups.

Furthermore, as experimenters want p values to reach at least .05, and .06 does not count, we predict that the records of psi-based performances should look like “scree jumps,” that is, sudden step-wise declines in scoring. Of course, file drawer problems and selective publication may also contribute to this effect and make it difficult to detect any true effect.

Another approach would be to use covert or semicovert psi tests to study the performance of successful researchers. Besides the obvious ethical issues, the long-term effects of such an undertaking would need to be considered. Alcock (2010) outright rejects the experimenter effect as a mere ad hoc explanation for when replications fail. Some of the material reviewed here might at best cause critics to pause, but as Alcock says, predictive evidence is called for. This seems to be a reasonable requirement, and one way of fulfilling it is extensive testing of experimenters prior to their performance with participants. Of course, a prerequisite is that such a study would have to be done under the leadership of a psi-conducive experimenter. The problem is that major studies of this kind are unlikely to be funded in the present climate, but they should be given priority.

The above review strongly suggests that psi abilities are not highly represented in either experimenters or in subject-participants as a group. Moreover, it seems likely that psi occurs in the context of the subtleties of human interaction, and only when we get the phenomena at a high level can we learn something fundamental about them; this means we will need to be highly selective in recruiting participants. A similar impasse has recently been reached in hypnosis research over the existence of a special (trance) state, leading to a refocusing on studying especially gifted individuals, single virtuosos rather than groups (Parker, 2014; Woody & Sadler, 2005).

At present, it seems likely that the term “experimenter psi” is seen as an assault on the basic tenet of objectivity in science (Kennedy, 2003). An encounter of the first author with a former fellow student, now a prominent researcher in cognitive psychology, made this clear when the former student remarked that parapsychology would be banned from his department at Cambridge University because, if psi exists, many of the secure findings that
ensure the status of psychology would be forthwith undermined.

This raises a larger concern, which Millar (2012) expressed when he wrote: “Another possibility is that psychological research is as contaminated by experimenter psi as parapsychology itself” (p. 19). We both agree that given the apparent pervasiveness of the experimenter psi effect in parapsychology, it almost certainly occurs in other sciences. There are even signs that the experimenter psi effect is bedeviling psychology and may eventually be noticed as an “anomaly” without parapsychologists needing to reveal their secret (Kahneman, 2012; Schooler, 2011; Editorial, 2013). Covert testing of researchers apparently possessing the so-called Midas Touch might contribute to this process. To avoid a backlash, such projects might euphemistically be called the study of “the King Midas Touch” or merely “self-efficacy amongst researchers.”

But let’s not avoid the inconvenient question: Do we continue to keep the secret or divulge it? Given how the university base of parapsychology in Europe and North America is being eroded and the funding of doctoral students no longer meets the high costs of university overheads, there seems little left to lose (Parker, 2013b). Seen in the long-term perspective, should experimenter psi eventually be accepted in other sciences, parapsychology would assume a new status. The discovery that experimenter effects are often psi-based would change our conception of science in ways difficult to imagine.

References


Honorton, C. (1976). Has science developed the competence to confront the claims of the paranormal? [Presidential Address]. In J. D.


Parker, A. (2012). The nearer your destination, the more you are slip slidin’ away. *Journal of Parapsychology, 76*, Supplement, 40–42.


Acknowledgement

We are grateful to the Bial Foundation for their support of this project.

Abstracts in Other Languages

German

DIE ENTHÜLLUNG VON PSI-GEHEIMNISSEN: DER ERFOLG ERFOLGREICHER EXPERIMENTATOREN SCHEINT SICH IHREM EIGENEN PSI ZU VERDANKEN

ZUSAMMENFASSUNG: Eine Literaturübersicht ergibt starke Anhaltspunkte dafür, dass erfolgreiche Experimen-
tatoren mittels ihrer eigenen Psi-Fähigkeit zu positiven Ergebnissen kommen – sei es direkt oder im Zusammen-
wirkung mit derjenigen ihrer Versuchsteilnehmer. Die Lebensgeschichten wohlbekannter Experimentatoren legen
nahe, dass in einigen Fällen diese Fähigkeit bereits vor ihrem Erfolg als Experimentatoren bestand, bei anderen
scheint sie sich im Laufe ihrer Experimente gezeigt zu haben oder diese wurden darauf aufmerksam. Es werden
Beispiele für das erfolgreiche Abschneiden psi-förderlicher Forscher gegeben, wenn sie selbst an Experimenten
teilgenommen haben. Einige bisher unveröffentlichte Aspekte der Gothenburger Ganzfeldstudien werden vorges-
estellt, die diese Schlussfolgerungen unterstützen. Theoretische Hintergründe im Sinne der gegenwärtigen Beo-
bachtetherorien und gegenwärtiger Auffassungen von Bewusstsein werden kurz vorgestellt. Auf welchen Wegen
sich der psi-bedingte Experimentatoreinfluss auf „forensische“ oder andere Art äußern kann, wird ebenfalls darg-
estellt. Dem Experimentatoren-Psi werden weitreichende Implikationen sowohl für Psychologie wie auch Parapsy-
chologie eingeräumt.

Spanish

DEVELANDO LOS SECRETOS DE PSI: LOS EXPERIMENTADORES CON ÉXITO USAN SU PROPIO PSI

RESUMEN: Una revisión de la literatura apoya fuertemente que los experimentadores con éxito utilizan su propio
psi para producir resultados positivos, ya sea directamente o de forma concertada con la de sus participantes. Las
historias personales de experimentadores conocidos sugieren que en algunos casos existía esta capacidad antes de
su éxito como experimentador, para otros parece que se ha producido o hecho consciente a través de su experi-
mentación. Se dan ejemplos de la participación exitosa de investigadores psi con éxito cuando eran participantes en
otros experimentos. Se presentan algunos aspectos inéditos de los estudios ganzfeld en Gotemburgo que apoyan
estas conclusiones. Se revisan someramente antecedentes teóricos en términos de la teoría actual de observación
actual y la visión contemporánea de la conciencia. También se presentan algunas formas de revelar la influencia del
experimentador psi “forensicamente” o de otra forma. El psi del experimentador psi tiene implicaciones de largo
alcance para la psicología y la parapsicología.

French

REVELER LES SECRETS DU PSI : LES EXPERIMENTATEURS A SUCCES SEMBLENT EMPLOYER LEUR PROPRE PSI

RESUME : Une revue de la littérature vient démontrer que les expérimentateurs les plus en réussite emploient
leur propre psi pour produire des résultats positifs, soit directement, soit de concert avec leurs participants. Les
histoires personnelles de fameux expérimentateurs suggèrent que, dans certains cas, cette capacité a préexister à leur succès en tant qu’expérimentateur, et pour d’autres elle semble avoir émergé ou avoir été reconnue au cours des expérimentations. Nous donnons des exemples de réussites de chercheurs facilitant le psi lorsqu’ils sont eux-mêmes participants à des expérimentations. Certains aspects inédits des études ganzfeld de Göteborg sont présentés à l’appui de ces conclusions. Les arrière-plans théoriques en termes de théorie observationnelle et autres modèles de la conscience sont brièvement passés en revue. Nous présentons aussi certaines manières de détecter ou de révéler l’influence de l’expérimentateur psi. La notion d’expérimentateur psi pourrait avoir d’importantes implications à la fois pour la psychologie et pour la parapsychologie.
ABSTRACT: We examine the critique of parapsychology offered by Professor Richard Wiseman in his 2010 paper, Heads I Win, Tails You Lose; How Parapsychologists Nullify Null Results, published in the Skeptical Inquirer, and offer detailed rebuttals to his main contentions. Some of the analyses we conduct are as follows: We compare reproducibility of psi experiments to reproducibility of experiments across related mainstream fields, finding that they are similar. Using both theoretical and empirical approaches, we demonstrate that file-drawer effects are not significant in the ganzfeld. We scrutinize and critique cases of alleged experimenter nullification of null results. We challenge—and offer alternatives to—the conclusions of the Milton and Wiseman meta-analysis, based on findings from Bem, Palmer, and Broughton, as well as our own results. We show that the evidence for ostensible declines in the actual effects of ganzfeld and forced-choice ESP paradigms is largely illusory and challenged by findings of recent inclines. Finally, we present strategies for progress according to the most compelling trends and consistencies we have found in the present database. These results, we hope, serve an illustrative purpose: a case examination of criticism in parapsychology with Wiseman as the main example, showing the degree to which the literature seems to support psi as the most plausible explanation of the data.

Keywords: Wiseman, ganzfeld, critique, skepticism, psi, parapsychology

Written in the spirit of the contributions made to Krippner and Friedman’s (2010) book, Debating Psychic Experience, we aim in this essay to contribute to the ongoing conversation on psi and science. Many reviews and meta-analyses have been published which examine the data, including very recent ones—our aim is to examine the criticism. For this purpose, we selected a well-known general critique of the field by Wiseman (2010a).

The arguments of that critique have not been extensively rebutted before. Carter (2010a), in “Heads I Lose, Tails You Win: How Richard Wiseman Nullifies Null Results and What To Do About It,” replied to Wiseman, but his rejoinder concentrated most heavily on Wiseman’s own conduct as an experimenter and not so much on his arguments. We address the latter to the best of our ability, and we keep our analysis manageable by placing special emphasis on the ganzfeld experiments, the “flagship” of parapsychology (Parker, 2000).

In the interest of full disclosure, our position is that these experiments and others have produced robust evidence for a communications anomaly of the type outlined by Bem and Honorton (1994)—though we reserve opinion on whether this is, ipso facto, psi—to such a degree that they necessitate analysis and replication from the mainstream. This is due both to careful precautions of investigators over the years as well as to surprising consistencies in the data, which we explore. Our paper ends with a point of agreement between us and Wiseman, illustrating the possibilities for future research.

If our comments and suggestions aid the development of parapsychology as a field, or conversely, the improvement of skeptical analysis, we will consider our job well done.

The Perception of Null Results

The major premise of Wiseman’s critique is that parapsychologists tend to accept positive results as evidence for psi but dismiss null results with post hoc explanations. In this regard, Wiseman writes:

Parapsychologists frequently create and test new experimental procedures in an attempt to produce laboratory evidence for psi. Most of these studies do not yield significant results. However, rather than being seen
Crucial to the strength of Wiseman’s critique is the question of how much weight null results should reasonably carry in the assessment of the evidence for psi—and what kind of null results are at issue. But before we address this, we note that although it is true that most studies in parapsychology databases do not display significant results, it is also true that the number that do is significantly above the null hypothesis expectation. Consider, for example, the post-PRL database, which consists of the studies in the Milton and Wiseman (1999) and Storm, Tressoldi, and Di Risio (2010) meta-analyses, covering the period 1988–2008. These 60 studies were conducted following a seminal report from Honorton’s Psychophysical Research Laboratories (PRL; Bem & Honorton, 1994), after the strict methodological guidelines proposed by Hyman and Honorton (1986). Only 15 of these post-PRL studies (25%) were significant at \( p \leq .05 \), whereas under the null, only 5% should have met this threshold, and the probability of getting 15 or more significant studies by chance alone is less than 1 in 5,200,000. Thus, average investigators have a probability of producing significant results that is five times what they would have if nothing significant was occurring in these experiments. We consider that important. Indeed, it is on this sort of observation that the ganzfeld, and similar domains of research, rest their claim to repeatability.

But is it sufficient? We note that there are several valid metrics by which to gauge reproducibility, and it is beyond the scope of this paper to present them all (see Cumming, 2012; Utts, 1991). The metric we focus on is the proportion of significant studies \( (p \leq .05) \) produced by a given research technique, a result governed by statistical power, or 1-\( \beta \). This can be thought of as the probability of obtaining significance given the attributes of one’s research methodology, and it is a direct function of type of significance test, effect size (ES), sample size (\( N \)), and alpha (\( \alpha \)) level. Because power governs the potential success of a study, we believe it critical to consider power before judging what level of reproducibility one should be seeing in a field as a litmus test of validity; after all, when power is low, we will fail to detect even a completely consistent effect more often than not.

In this vein, it is reasonable to ask how much power is employed in parapsychology, generally. According to Utts (1991) and Tressoldi (2012), not a lot. Taking the ganzfeld as a prototypical example, for the 105 studies reported in Storm, Tressoldi, and Di Risio (2010) with four-choice designs, the overall hit rate is 32.2% and the mean sample size is 42, for an average power per study of about 30%; this value comes close to the proportion of significant studies (28.5%) in that sample. Similar calculations performed by Derakhshani (2014)—using his own power test and one recommended by Ioannidis and Trikalinos (2007)—demonstrate that the proportion of significant studies in all past ganzfeld databases can be accurately predicted using standard power assumptions, within 95% confidence intervals. This suggests that ganzfeld studies elicit the level of consistency that is expected given the characteristics of those studies, and that they are replicable insofar as we can make predictions about their probability of success and have them verified. The evidence is that psi effects, at least in the ganzfeld, lawfully follow the predictions of conventional statistical models to a degree that is conducive to scientific investigation.

We should thus be able to reliably effect changes in our levels of success, using these models. If we aim for 80% power in the ganzfeld, for example, we may try increasing sample size alone; however, this will result in at least 236 needed trials (given the 32.2% hit rate found in Storm et al., 2010)—a quantity likely to be inaccessible to the average investigator. In fact, the largest number of trials ever run in a single ganzfeld study is 138 (Parra & Villanueva, 2006). Another option to boost power is to raise the ES of studies. Derakhshani (2014) takes this route and shows, based on the post-PRL database, that if investigators use only selected participants (e.g., participants with prior psi experience, mental discipline practice, prior psi training, belief in psi, or preferably a combination of these)—a population that achieved a 40.1% hit rate in the post-PRL database—they would need only 56 trials for 80% power. We note that this predicted higher proportion of significant studies is not only completely consistent with past findings, but practicably attainable.

Another question we might ask about power and replication in parapsychology is how they stack up with what is found in other sciences. To our knowledge, there has never been an in-depth comparison of this type, but one is sorely needed. For example, in Richard, Bond, and Stokes-Zoota’s (2003) exhaustive meta-analysis of 322 meta-analyses in social psychology, the average statistical power was 20%, a little below that of the post-PRL database. With this power, the typical social science experiment would need at least 173 trials to achieve 80% reproducibility (at \( p \leq .05 \)), which is already considerably higher than normal (Hartshorne & Schachner, 2012). The reason for this is that ESs in social psychology are usually small—about \( r = .21 \) on average—and researchers tend not to conduct
large enough studies to compensate for this. In fact, almost a third of the ESs reported in Richard et al. (2003) were $r = .1$ or below, requiring an average $N$ of 772 just to achieve a power of 80% (Hartshorne & Schachner, 2012).

Hartshorne and Schachner (2012) write, additionally, that according to multiple meta-analyses, the statistical power of a typical psychology or neuroscience study to detect a medium-sized effect (defined variously as $r = .3$, $r = .4$, or $d = .5$) is approximately .5 or below (Bezeau & Graves, 2001; Cohen, 1962; Kosciulek & Szymanski, 1993; Sedlmeier & Gigerenzer, 1989). (p. 2)

But in fact, for small effects ($d \leq 3$), this power is much lower. Rossi (1990) observed a mean power of 17% across 221 articles for ESs in this range, in three prominent psychology journals starting in 1982. Neuroscience research has also been recently reviewed by Button et al. (2013), who looked at 730 studies in 49 meta-analyses and concluded that the median statistical power for that discipline was about 21%. They subsequently observed that the removal of seven outlying meta-analyses with very large effect sizes brought their power estimate to 18%. All of these power values—from the average power in social psychology, the mean power for small effects in psychology, and the median power for neuroscience studies—fail to meet the average power for a ganzfeld study conservatively calculated at 30%, for all 105 studies in Storm et al. (2010). Considering just the recently gathered 30 ganzfeld studies from 1997 to 2008 (Storm et al., 2010), the average power is actually much higher, at approximately 43%. Even for all the nonganzfeld free-response studies reported during that period in Storm et al. (2010), the mean power of 19% (excluding four studies not of four-choice design) is still marginally greater than for most of the aforementioned mainstream areas.

Bakker, Van Dijik, and Wicherts (2012) estimate, moreover, that for the average ES of psychology research ($d = 0.5$, which they note is skewed by publication bias), using a two independent samples comparison, the power of psychology studies across multiple meta-analyses is about 35% (p. 544). Despite the roughness of this estimate, it happens to closely match the reported current proportion of significant results in the Reproducibility Project database (33.3%), a meta-experiment with a median power of 95% to detect effects across a wide range of replications of papers representatively sampled from psychology journals (Nosek, Lai, LeBel, Gilbert, & Strohminger, 2014). Why are these percentages so similar? The answer is that publication bias in psychology is very prevalent, so that if we assume a simplified model reasonably close to the truth, all published psychology studies are significant. For psychology studies with true effects, then, following Derakhshani (2014) and Ioannidis and Trikalinos (2007), our mean power estimate says that 35% will reach significance and get published. Therefore 35% should very roughly be the proportion of significant published studies with true effects. The other 65% should be false positives drawn from studies with no true effects. So when experiments such as those in the Reproducibility Project, using extremely high power, representatively replicate from all published significant studies, those 35% of studies with true effects should be the studies that are successfully replicated. Since this seems to be the case, it confirms the predictions of Derakhshani (2014) and Ioannidis and Trikalinos (2007) that, in the presence of a consistent effect, the average power in a field should serve as a good quantifier of reproducibility, per our definition.

On this subject, Nosek (2012) writes:

There exists very little evidence to provide reproducibility estimates for scientific fields, though some empirically informed estimates are disquieting (Ioannidis, 2005). When independent researchers tried to replicate dozens of important studies on cancer, women’s health, and cardiovascular disease, only 25% of their replication studies confirmed the original result (Prinz, Schlange, & Asadullah, 2011). In a similar investigation, Begley and Ellis (2012) reported a meager 11% replication rate. (p. 657)

In the face of these reproducibility estimates, we argue that for any area of parapsychology to achieve a replication rate of 25% to 30% to 37%—the proportion of significant results in the post-PRL, the whole ganzfeld, and the most recent 30 studies, respectively (Storm et al., 2010)—which we have shown to be comparable to other sciences; is in fact quite remarkable, given that the total human and financial resources devoted to psi research from 1882 to 1993 has been estimated to comprise less than two months’ research in conventional psychology (Schouten, 1993, p. 316). This observation warrants the conclusion that not only is the ganzfeld technique consistent, but it is also progressing at a rate similar to that of mainstream social and behavioral fields—and surprisingly so, given its
resources. The conformance of the ganzfeld database to power predictions, moreover, strongly suggests that adoption of strategies to boost power would improve reproducibility, and that attempting to do so would be a worthwhile venture.

**Investigating the File Drawer**

For a meta-analysis to be valid, arguably the most important criterion is that all of the data are there to analyze—or at least that no systematic bias of any importance is present in the studies selected. Yet this is what Wiseman (2010a) seems to imply by his comments:

Once in a while one of these [parapsychology] studies produces significant results. Such studies frequently contain potential methodological artifacts, in part because they are using new procedures that have yet to be scrutinized by the research community ... the evidential status of these positive findings is problematic to judge because they have emerged from a mass of nonsignificant studies. Nevertheless, they are more likely than nonsignificant studies to be presented at a conference or published in a journal. (p. 37)

Firstly, it is important to note that the idea that positive studies are more likely to contain methodological artifacts is poorly supported for research into ESP (though it does receive some support for recent research into psychokinesis, as seen below). We are aware of one meta-analysis by Schmidt, Schneider, Utts, and Walach (2004) that found a significant negative correlation between overall quality and ES for direct mental interaction with living systems (DMILS) studies, but not remote staring studies. These correlations are rare. Storm et al. (2010) showed, for example, that for their free response studies conducted from 1992–2008, quality ratings obtained under blind conditions did not correlate significantly with ESs: \( r(65) = .08, p = .11 \). We were further able to demonstrate that for groups of high-scoring selected participants in Storm et al.’s (2010) 30-study ganzfeld database, the mean study quality rating was greater than for the significantly lower-scoring unselected participants (\( q = 0.84 \) and 0.79, respectively, where \( q = 1.00 \) was the highest possible rating). We give a sampling of the literature on the question of quality-ES correlations as follows, endeavoring to use only the most recent results for each paradigm of research:

1. The only meta-analysis of physiological presentiment studies conducted to date detected a nonsignificant positive correlation between methodological stringency and ES: \( r = .21, 95\% \text{ CI} = -.20–.53 \) (Mossbridge, Tressoldi, & Utts, 2012).

2. A meta-analysis of forced-choice precognition studies yielded a very small and nonsignificant positive correlation between ES and study quality; \( r = .08, p = .20, \text{ two-tailed} \) (Honorton & Ferrari, 1989).

3. In a review of the success of the forced-choice ESP paradigm in parapsychology, a very small and nonsignificant negative relationship was found between ES and quality ratings, and thus no dependency; \( r = -.08, p = .48, \text{ two-tailed} \) (Storm, Tressoldi, & Di Risio, 2012).

4. Bösch et al. (2006) found a highly significant correlation between ES and safeguard sum score in their database of RNG studies, indicating that lower quality studies produced larger ESs: \( r(386) = .15, p = .004 \). They noted, however, that the average quality of these studies was very high.

In view of these considerations, the hypothesis that experimental flaws are systematically and inversely related to study ES in parapsychology should be seen as generally unsupported by the evidence, unless analyses using novel quality ratings find conflicting results.

Wiseman’s main criticism, however, raises a concern that parapsychologists have been conscious of for decades: the file-drawer problem. Its premise is that studies with positive results are more likely to find their way into meta-analytic databases than studies with negative results, and that this therefore creates a systematically biased sample. This effect has been well-documented (Ahmed, Sutton, & Riley, 2012; Fanelli, 2010; Rothstein, Sutton, & Bornstein, 2005). Fanelli (2010), for example, observed that 84% of publications in various sciences reported positive results—a very unlikely proportion given the low power estimates discussed in the previous section of
this paper—with psychology reporting the most: 91.5%. This estimate for psychology is only minimally different from previous values reported by Sterling (1959) and Sterling, Rosenbaum, and Weinikam (1995), at 97% and 96% respectively. It is common practice for journals to reject null studies in favor of positive ones—the result being that many unsuccessful studies never make it to publication, and thus escape detection by meta-analysts. Even if a study does get into print, it may still be excluded from meta-analytic consideration; biases inherent in the meta-analytic search process or inclusion criteria may cause the study either to be overlooked or disregarded. We make a distinction between these two types of selection bias, calling the first publication bias and the second inclusion bias (although both are problematic, the former is arguably more so, as unpublished studies are less likely to be found than published studies).

Based on these reasons, then, we note that the selection bias criticism is a priori an extremely powerful one, but as we hope to show for parapsychology, ultimately untenable. One reason is that awareness of the file drawer came early for psi researchers. The earliest systematic cross-laboratories meta-analysis in scientific history, reported in Extra-Sensory Perception After Sixty Years (Rhine, Pratt, Stuart, Smith, & Greenwood, 1967), included a statistical method to estimate the influence of publication bias. Additionally, in 1975, the Parapsychological Association (PA) became the first scientific organization to adopt an official policy of publishing null results (Carter, 2010a). Beyond explicitly minimizing the file drawer, this decision brought into common psi research practice techniques designed to measure study selection bias, such as funnel plots, Rosenthal’s fail-safe N, and trim-and-fill methods, all of which have been used in reviews of psi research to argue effectively against the file-drawer explanation.

With regard to the ganzfeld, for example, Storm et al. (2010) applied Rosenthal’s fail-safe N (Harris & Rosenthal, 1985, p. 189) and found that no fewer than 2,414 unpublished studies with overall null results (i.e., \( z = 0 \)) would have to exist to reduce their 108 ganzfeld study database to nonsignificance. This is not a likely scenario. However, some have argued that Rosenthal’s calculation overestimates the file drawer (Scargle, 2000) by definition, because it implicitly assumes the reservoir of unpublished studies to be unbiased (\( z = 0 \)) instead of directionally negative (\( z < 0 \)). To overcome this problem, there are more conservative procedures such as the Darlington and Hayes (2003) method, which allows for a large proportion of unpublished studies to have negative \( z \) scores. Applying this method as an additional check for the same homogeneous 102-study database, Storm et al. (2010) showed that the number of unpublished studies necessary to nullify just their 27 studies with statistically significant positive outcomes was 384, and 357 of these could have \( z < 0 \). Given the official policy of publishing null results set down by the PA, and the small number of scientists conducting research in this area, such a large number of negative studies can only be deemed highly untenable.

With regard to the validity of Rosenthal’s fail-safe \( N \), we agree with the technical correction put forward by Scargle (2000) that the theoretical mean \( z \) of unpublished studies for an extreme file-drawer case, under a null distribution, is -0.1085, not 0. Harris and Rosenthal (1988) note, however, that “Based on experience with meta-analyses in other domains of research (e.g., interpersonal expectancy effects) the mean \( z \) or effect size for nonsignificant studies is not zero but a value pulled strongly from zero toward the mean \( z \) or mean effect size of the obtained studies (Rosenthal & Rubin, 1978)” (p. 45). Their assumption that the average \( z \) score of excluded studies is zero is therefore a conservative one for most any distribution that is shifted some positive distance from a null distribution, and although this specifically indicates situations where an effect is present, we argue that the evidence for such an effect in the ESP literature is overwhelming, whatever one may believe about its underlying cause. Another conservative assumption in Rosenthal’s procedure is that each excluded study is considered to have a sample size equal to the average sample size of the meta-analysis, whereas overlooked studies tend to be smaller.

Further evidence against the file-drawer effect in the ganzfeld, supporting the notion that unpublished studies show directionally positive results, comes from a mail survey by Blackmore (1980), who queried parapsychologists conducting ganzfeld experiments to obtain a direct estimate of the file drawer. The returned questionnaires revealed 32 unreported studies, 12 of which were still in progress, and one that could not be analyzed. Of the 19 remaining, 14 were judged to have adequate methodology, including 5 that were significant (36% of the total). This proportion of significant results is statistically unlikely according to the null hypothesis; in fact, it yields an exact binomial result of \( p = .0004 \), or odds against chance of 2,342 to 1. So the file drawer itself is—directly counter to the skeptical prediction—inclined towards the psi hypothesis. Furthermore, the proportion of significant studies in Blackmore’s 1980 paper (5 out of 14, or 36%) is not significantly different from the proportion found in Honorton’s 1985 meta-analysis (12 out of 28, or 43%), Fisher’s exact \( p = .46 \), one tailed. Given this information, it is not surprising that Blackmore (1980) concluded that “the bias introduced by selective reporting of ESP ganzfeld studies is
Beyond the Coin Toss

not a major contributor to the overall proportion of significant results” (p. 217). Blackmore’s survey must be understood in context, however; it took place more than 34 years ago, and 20 studies in it were destined for publication. As such, it can only be considered a snapshot of the file drawer at a given time.

Additionally, even if one entertains the notion that the included ganzfeld studies are drawn from an overall statistically null distribution—in spite of the results of the conservative Darlington-Hayes calculation and the Blackmore (1980) survey—the parapsychological practice of considering significantly negative results to be “psi-missing,” and therefore potential evidence for psi, helps to ensure that the negative tail of this distribution is also included, meaning that the average $z$ of the excluded studies should be relatively close to zero, not highly negative. This symmetrical exclusion principle is supported by Harris and Rosenthal’s (1988) assessment of the ganzfeld, which yielded evidence consistent with “larger positive and larger negative effect sizes than would be reasonable” (Harris & Rosenthal, 1980, p. 44), although by a small margin.

Perhaps most persuasively, as we showed in the first section of this paper, the average power of ganzfeld studies across databases accurately predicts their proportion of significant results, suggesting minimal or no selection bias (Ioannidis & Trikalinos, 2007). Similar calculations to Rosenthal’s and Darlington and Hayes’, as well as funnel plots and trim and fill algorithms, have plausibly written the file-drawer explanation out of other paradigms in parapsychology, including remote viewing studies (Tressoldi, 2011), psychokinesis studies (Radin et al., 2006), forced-choice ESP studies (Tressoldi, 2011), and precognition studies (Honorton & Ferrari, 1989). Collectively, they provide evidence that selective reporting is not a significant factor in psi research.

There is, however, a still more direct way to tackle Wiseman’s (2010a) criticism, since in his words “... only one paper has revealed an insight into the potential scale of [the file-drawer] problem”(p. 37). That paper is the Watt (2007) Koestler Parapsychology Unit report, which surveyed all parapsychology undergraduate projects undertaken and supervised by the Edinburgh staff between 1987 and 2007. About it, Wiseman (2010a) says:

Only seven of the 38 studies had made it into the public domain, presented as papers at conferences held by the Parapsychological Association ... there was a strong tendency for parapsychologists to make public those studies that had obtained positive findings, with just over 70 percent (five out of seven) of the studies presented at conferences showing an overall significant result, versus just 15% (3 out of 20) of those that remained unreported. (p. 37)

At first glance, this appears to be incontestable proof of a serious publication bias, but a closer look at what Wiseman says is instructive. First, the very fact that meta-analyses in parapsychology include studies presented at conferences but not published in journals (an uncommon practice in the sciences) testifies to its attempt to combat selective reporting (note that PA conference papers are still peer reviewed). Second, Wiseman makes a critical mistake when he mixes projects as varied as “dowsing for a hidden penny, the psychokinetic control of a visual display of a balloon being driven by a fan onto spikes, presentiment of photographs depicting emotional facial expressions, detecting the emotional state of a sender in a telepathy experiment, ganzfeld studies, and card guessing” (p. 37) and then gives the inflated 70% and 15% figures as evidence for a massive file-drawer effect. Because these studies fall into different experimental paradigms, and some of them do not belong clearly to any defined line of research (i.e., they are purely exploratory), mixing them together tells us nothing about the evidential impact of this file drawer on proof-oriented meta-analyses.

It can be seen, for example, that if just one type of study is taken from Edinburgh’s varied selection—ganzfeld studies—Wiseman’s criticism is rendered moot. Of the 38 KPU undergraduate projects that tested for a psi effect, only 5 were ganzfeld (one by Colyer and Morris, cited by Watt, 2006; one by Morris, Cunningham, McAlpine, and Taylor, 1993; two by Morris, Summers, and Yim, 2003; and one by Symmons and Morris, 1997). Furthermore, although the nonsignificant Colyer and Morris study was the only study not presented at PA conventions, the Morris et al. (1993) study was presented, and was also nonsignificant. This leaves a single study in the file drawer whose reasons for not being included are unknown, and whose exclusion is not enough to say anything meaningful about selective reporting in the ganzfeld.

Putting aside ganzfeld studies, three additional student projects were presented at the PA conventions, and they were all DMILS studies. Two had significant results (Brady & Morris, 1997; Delanoy & Sah, 1994) and one was nonsignificant (Watt, Hopkinson, & Fraser, 2006). Examination of Watt (2007) revealed that five of these studies—by Howat et al., Juniper et al., Martin and Miller, Phillips and Morris, and Robert et al.—were not presented at
None was statistically significant; however, two were quite low-powered with only 28 participants each, making statistical significance difficult to achieve. All three found effects in the predicted direction and of a magnitude ($r = .15$) larger than that found in the Schmidt, Schneider, Utts, and Walach (2004) meta-analysis of 15 remote staring studies (with the latter finding a mean effect size $d = 0.13$, which converts to $r = .079$). However, the two databases are not independent because Schmidt et al. retrieved unpublished studies, including two of the three studies reported here (Howat, and Juniper & Edlmann). (p.348)

Furthermore, Schmidt et al. (2004) made an effort to locate unpublished experiments; they contacted authors of all the published studies to ask for assistance and posted a search request on an e-mail forum discussing parapsychological research issues. Watt’s DMILS studies are therefore unlikely to serve as evidence of a significant file drawer. The KPU ganzfeld pool, however—because of Watt’s comprehensive survey—is an example of a dataset that we can reasonably infer possesses no selective reporting of studies. If we consider the five studies provided, including the Colyer and Morris study, for a total of 195 trials and a hit rate of 33.8%, the cumulative probability of their results under the null hypothesis is $p = .004$ (one-tailed, exact binomial). The 10-study PRL database, too, is known to have no selective reporting: Bem and Honorton (1994) explicitly stated that “the 11 studies just described comprise all sessions conducted during the 6.5 years of the program. There is no file drawer of unreported sessions” (p.10) (Note: it is common in analyses of the PRL studies that one highly successful study, Study 302, is removed from analysis due to well-known concerns about optional stopping, thereby leaving 10 studies). Additionally, Honorton (1985) states, “Except for two pilot studies, the number of participants and trials was specified in advance for each series. The pilot or formal status of each series was similarly specified in advance and recorded on disk before beginning the series. We have reported all trials, including pilot and ongoing series, using the digital autoganzfeld system. Thus, there is no ‘file-drawer’ problem in this database” (p. 133). This file drawer free database has a hit rate of 32.2%, 329 trials, and a binomial probability of $p = .002$. Given that these hit rates are not significantly different from each other, we can merge the two datasets to form one 15-study pool with no file drawer, 524 trials, a hit rate of 32.8%, and a binomial probability of $p = 5.91 \times 10^{-8}$. This composite hit rate (32.8%) is close to that of the remaining 90 studies in Storm et al.’s (2010) database. When we remove these 15 studies, as well as 3 not of four-choice design, there remain a total of 3,516 trials with a composite hit rate of 31.8%. This convergence of results from three analyzed study pools (KPU, PRL ganzfeld, and the rest of the ganzfeld studies in the Storm et al. database) suggests that if there is a contribution from selective reporting to the overall hit rate, it is likely to be negligible or nonexistent. It is also an example of a surprising consistency in psi research.

In sum, although we acknowledge that we cannot comment as extensively on other paradigms of parapsychology as we can on the ganzfeld, at present we believe that the ganzfeld has performed admirably with regard to the file drawer. If this protocol can be considered representative of parapsychology as a whole, selective reporting of positive results cannot be considered to have significantly influenced the evidence for the existence of psi phenomena.

**Parapsychology and Null Results**

Wiseman’s (2010a) next major criticism involves variations in the procedure of parapsychology experiments:

If a procedure seems to yield significant psi effects, additional follow-up studies using that procedure are conducted. Although these additional studies occasionally take the form of strict replications, they usually involve some form of variation. If these follow-up studies obtain significant results, they are often the subject of considerable debate: proponents argue that the findings represent evidence of psi, and skeptics scrutinize the work for possible methodological and statistical shortcomings. However, any failure to replicate can be attributed to the procedural modifications rather than to the nonexistence of psi. (p.37)
Although Wiseman’s critique in some respects makes a legitimate point, it should be remembered that counter-advocates spend much of their time doing just what Wiseman opposes, but in reverse, and this has been well documented (Carter, 2010a). It is thus important to analyze instances of claimed spurious nullification to determine whether they represent (as advocates believe) genuine attempts to understand a phenomenon, or (as counter-advocates believe) a simple dismissal of what would otherwise be considered a failure. Wiseman provides two cases for us to examine.

His first piece of evidence for retrospective nullification of null results is a paper by Kanthamani and Broughton (KB; 1994), which reported an attempt to replicate the ganzfeld effect that yielded null results. Wiseman criticizes them for making no mention of the null hypothesis as an explanation for their nonsignificant findings, instead concluding that “it is probably safe to say that static picture targets remain a less than ideal choice for ganzfeld experiments” (Wiseman, 2010b). It is clear that Wiseman is implying that this decision was arbitrary and unwarranted, but evidence from the paper and from previous analyses contradicts his conclusion. Bem and Honorton (1994) report, for example, that among the 28 pre-PRL studies, 9 used “dynamic” targets (View Master slide reels) as compared to static pictures, and those 9 found a significantly higher hit rate than the other 19 (50% vs. 34%, respectively; Fisher’s exact $p = .04$, two-tailed). Honorton’s own PRL studies (Bem & Honorton, 1994) compared 164 dynamic targets to 165 static targets and also found a significant difference in scoring rates (37% vs. 27%, respectively; Fisher’s exact $p < .04$). Therefore, when KB found a 27.6% hit rate in the 350 trials for their static targets, they wrote that they had replicated the finding by Bem and Honorton—and they had, to a very precise degree.

KB also found that the four groups of participants in their database that conformed to one of the four measures of “optimal subjects” as defined by Honorton (1997)—previous psi experiences, previous psi testing, a feeling-perception (FP) personality on the Myers-Briggs Type Inventory, and practice of a mental discipline—produced overall hit rates ranging from 31% to 36%. This finding is of significant importance considering that this same subpopulation aggregate for the PRL and FNRM databases—the latter an independent replication of the PRL trials (Broughton, Kanthamani, & Khilji, 1989)—was 31% (Honorton, 1997). Moreover, when three of these optimal-participant measures were combined in the KB studies, forming what Honorton (1997) called the “three-predictor model,” the results were striking: KB’s database exhibited a hit rate of 41.3% (46 trials; exact binomial $p = .011$, one-tailed), whereas the PRL and FNRM databases yielded a combined rate of 43% (99 trials; exact binomial $p = .0004$, one-tailed). It should be noted that these results are surprisingly consistent, and not post hoc data selection; Honorton and Schechter (1987) originally found these predictors in the PRL-1 novice series before Honorton (1997) applied them to the PRL-2 novice series, as well as the independent FNRM database, shortly before his passing. Honorton (1997) wrote:

At the 1986 PA Convention, Honorton and Schechter (1987) presented an exploratory analysis of performance correlates for the first two PRL novice series (Series 101-102; hereafter designated PRL-1), suggesting that initial ganzfeld ESP performance was positively and significantly related to self-reports of personal psi experiences, Feeling/Perception (FP) preferences on the MBTI, and prior participation in nonganzfeld psi experiments. A positive but nonsignificant tendency for better performance among participants reporting involvement with mental disciplines such as meditation was also found. . . . In this paper, the PRL-1 findings will be compared with those in the later PRL novices series (Series 103-105; hereafter designated PRL-2) and the FRNM series to estimate the overall magnitude and consistency of the four predictors. (p. 143)

Here we should note that Honorton produced a “three-predictor model” in addition to his four-predictor model; the former was created because of the small number of subjects satisfying the prior psi testing condition, and omitted this requirement.

Recall now the results that KB found for their three-predictor dataset; if these are added to the total PRL and FNRM databases, there are 145 trials which yield a 42.06% overall hit rate (exact binomial $p = 5.07 \times 10^{-5}$, one-tailed). As for the omitted characteristic, Kanthamani and Broughton (1994) stated that prior psi testing was also successful, but because of the broader scope of the three-predictor model, they chose to apply it instead. This rather strongly confirms the improved performance of the selected participants, and it provides corroboratory evidence against the null hypothesis—even in light of the fact that the KB database overall is nonsignificant.

Because of these considerations, we argue that Kanthamani and Broughton (1994) were fully justified in noting that their studies confirmed the “PRL success model” (p. 7) in such a way that their conclusions cannot
be seen as evidence of retrospective nullification. Additionally, their results were not excluded from any relevant meta-analyses (Bem, Palmer, & Broughton, 2001; Milton & Wiseman, 1999; Storm et al. 2010), so even if their conclusions had been little more than confirmation bias, that would have had no effect on the evidence.

Wiseman’s next example of retrospective nullification mentions Melvyn Willin’s (1996a) study with musical targets as a prime model for “data mining,” which Wiseman defines as the tendency to search in the results of a null study for any correlation that can yield anomalous findings. Wiseman (2010a) criticizes Willin for his decision not to invoke the null hypothesis as an explanation for his failures:

Willin conducted a series of post hoc analyses, exploring, for example, the relationship between participants’ psi scores and their age, profession, hobbies, previous paranormal experiences, and relationship with the person acting as the sender. Additional analyses explored psi scoring as a function of the month and time of day each trial was conducted. Most of these analyses yielded inconclusive results, but Willin eventually found that trials conducted early in the experiment obtained a higher hit rate than those conducted later and suggested that this might have been due to “less interest being shown by the Receivers and the Senders or by an unintentional goat effect being displayed by the Experimenter.” (p. 38)

To counterbalance this information, it should be noted that Willin (1996a) began his paper with the sentence “experiments using actual music as the target have not been conducted very often” (p. 1), afterward listing several small exploratory attempts to elicit psi from musical targets, with mixed success (Altom & Braud, 1976; George, 1948; Keil, 1965; Shulman, 1938). This suggests that Willin knew his study was of a more exploratory than confirmatory nature; looking for trends and patterns after the fact was thus part of its design—especially given that it was the first large-scale experiment to employ musical targets (Parra & Villanueva, 2004). He even collected extensive background and personality data on his participants pre-analysis, for that precise purpose (Willin, 2005). We believe that post hoc findings are essential to science, so long as they are not counted as confirmatory (and the proper corrections for multiple analysis are applied), so we see no problem with his strategy. As for Willin’s attitude towards his null results, we suggest that the following comment from his follow-up study (Willin, 1996b), using previously high-scoring unselected participants, should be considered: “A chance hit rate of 25% was expected and a hit rate of exactly 25% was achieved. . . . These results thus provide no evidence for the communication of music by ESP” (Willin, 1996b, p. 103).

Although we do not doubt that there are instances of confirmation bias in the parapsychology literature, wherein researchers have perhaps given undue emphasis to a success while marginalizing a failure, our review of the two situations presented by Wiseman suggests the need to be critical of such claims when they do arise, as prospective examples of confirmation bias are themselves susceptible to confirmation bias.

Expectancy Effects in Parapsychology

In addition to his critique of Kanthamani and Broughton (1994) and Willin (1996a), Wiseman (2010a) briefly mentions experimenter effects: “Perhaps the most far-reaching version of this ‘get out of a null effect free’ card involves an appeal to the ‘experimenter effect,’ wherein any negative findings are attributed to the psi-inhibitory nature of the parapsychologist running the study” (p. 37).

Since Wiseman does not give readers much information about these effects, one is left with the impression that something is fundamentally wrong with parapsychology. But Wiseman does not mention that experimenter expectancy effects have been the subject of widespread research outside parapsychology (Rosenthal, 1976). It has been known for decades that subtle psychological variables such as Rosenthal and Jacobson’s (1992) “Pygmalion effect,” for example, can strongly impact participant performance and affect research in nearly all behavioral fields.

Given this information, it should not be surprising that the same effects occur in parapsychology. There is a long history of studying them in the literature (Smith, 2003). Wiseman himself was party to an experiment that tested the idea and found evidence for it (Wiseman & Schlitz, 1997). The study was of the “psychic staring” effect, in which half of the trials were conducted with Wiseman (a purportedly psi-inhibitory experimenter) as the experimenter/starer, and half with Marilyn Schlitz (a purportedly psi-facilitatory experimenter) in that role. Results were as predicted for the first and second collaboration: Wiseman found nothing and Schlitz found a small but statistically significant difference between the stare and no stare conditions.
Regarding this affair, journalist Guy Lyon Playfair (2014) wrote:

In the October 2002 issue of *The Paranormal Review*, Caroline Watt asked each of them [Wiseman and Schlitz] what kind of preparations they make before starting an experiment. Their answers were: Schlitz “. . . I tell people that there is background research that’s been done already that suggests this works . . . . I give them a very positive expectation of outcome. Wiseman: “In terms of preparing myself for the session, absolutely nothing.”

It does not help Wiseman’s case that he wrote, after the fact, that the testing process was “an enormously boring experience” (Watt, Wiseman, & Schlitz, 2002, p. 21) and that in most of the trials he was “pretty passive about it” (p. 22). However, the last collaboration failed to detect an effect with either experimenter (Schlitz, Wiseman, Watt, & Radin, 2006). At this point, both Schlitz and Wiseman reported feeling “burnt out” with the project, which, according to the experimenter expectancy effect hypothesis, could have led to a reduced performance for both Schlitz’s and Wiseman’s participants. Nevertheless, this interpretation should be taken cautiously, as it is retrospective. Of more interest is the fact that around 67% (two out of three) of Schlitz’s studies achieved significant effects, while 0% (none out of three) of Wiseman’s did. The contrast between their results becomes even more noticeable when the previous studies of these researchers (Wiseman et al., 1995; Wiseman & Smith, 1994; Schlitz & LaBerge, 1994) are included; Schlitz obtained a 75% (four out of five) success rate and Wiseman still had a 0% (none out of five) success rate.

Another interesting piece of evidence for a real and relevant effect comes from Smith’s (2003) discussion of a study by Judith Taddonio (1976):

Taddonio (1976) told student experimenters that the ESP test they were to use was a recently developed technique developed by Taddonio’s colleagues and that the students were being asked to conduct a replication of their findings. Taddonio manipulated the expectancy of experimenters by telling those in one group that participants in previous studies using this new technique had consistently obtained above chance scores. These experimenters were assured that the test could not fail and that the results of the student’s replication would give the same high scores. Experimenters in the second group were told that Taddonio’s colleagues who had developed the test were worried about it because participants were all scoring well below chance. They were led to believe that the test seemed to elicit psi-missing rather than psi-hitting and that there was no doubt that the student’s replication would show the same level of low scoring. In both a pilot study and a confirmatory study, participants tested by the experimenters given the positive expectancy about the test scored significantly higher than participants tested by the experimenters given the negative expectancy. (Smith, 2003, p. 75)

Taddonio’s results seem to support an explanation based, again, on experimenter-participant interaction; as she wrote, a “difference in the psychological impact of the two experimenter groups upon their subjects” (Taddonio, 1976, p. 113). Confirmation bias alone seems unable to account for these experimental differences.

On a similar note, regarding the possibility that experimenter effects result from motivated scoring errors, unconscious presence of flaws, and other similar explanations, there is good news for the ganzfeld. As Robert Rosenthal wrote in 2009 (cited by Carter, 2010b):

Ganzfeld research would do very well in head-to-head comparisons with mainstream research. The experimenter-derived artifacts described in my 1966 (enlarged edition 1976) book *Experimenter Effects in Behavioral Research* were better dealt with by ganzfeld researchers than by many researchers in more traditional domains. (p. 95)

Recent evidence supports this assertion for ESP research. Storm et al. (2010) compared the effect sizes obtained by different experimenters/laboratories in their 45-study ganzfeld and nonganzfeld noise reduction database, uncovering no significant differences between the groups using a standard ANOVA analysis (p = .32). For forced-choice studies, Storm et al. (2012) used the same analysis and found no evidence of a difference (p = .36). Mossbridge et al.’s (2012) presentiment meta-analysis, on the other hand, included no such comparison, probably...
because the level of homogeneity in their database ($I^2 = 27.4$) obviated the necessity for one. Based on the most up-to-date meta-analytic data, then, it is by no means clear that experimenter effects present a visible problem for psi research.

In a further study of experimenter effects, Watt and Nagtegaal (2004) found that, among all the sciences examined, parapsychology had taken the strongest precautions against experimenter effects by conducting 79.1% of its research using a double-blind methodology (compared to 0.5% in the physical sciences, 2.4% in the biological sciences, 36.8% in the medical sciences, and 14.5% in the psychological sciences). These findings are consistent with those of an earlier survey on experimenter effects by Sheldrake (1998).

To conclude, although it is recognized that experimenter effects may influence parapsychological results, it should now be seen that their occurrence in the field is not unique. Neither is the presence of these effects enough to dismiss the validity of the research, for whether an experimenter is psi-inhibitory or psi-conducive, ultimately all parapsychology studies will be included in attempting to draw conclusions about psi. We think the issue is a good bit more subtle than this. Such effects are, rather, inevitably involved in the process of studying and understanding a phenomenon whose properties are not fully known. Through continued research we may yet get a better grasp of them.

The Milton and Wiseman Meta-Analysis

Wiseman’s next criticism invokes the Milton and Wiseman (MW; 1999) meta-analysis, which found null results for psi across all post-PRL studies conducted until 1997 and spurred a significant debate in the parapsychology community about replication (Schmeidler & Edge, 1999). Before beginning to dissect its conclusions and methodology, however, we note that, as of the most recent meta-analysis (Storm et al., 2010), the overall hit rate of the post-PRL database remains highly significant. With this in mind, it is possible to visually gauge the impact of MW’s analysis by examining a plot of $z$ scores across the ganzfeld, including the MW dataset (Figure 1).
Beyond the Coin Toss

analysis (e.g., Bem et al., 2001; Honorton, 1985; Storm et al., 2010), although Storm et al. did present the binomial test as a supplementary analysis for their 30-study database from 1997–2008, and found highly significant overall outcomes for the 29 studies using a four-choice design. Nevertheless, the Bem and Honorton (1994) meta-analysis, which Milton and Wiseman were attempting to replicate, obtained its highly significant overall hit rate using the binomial test. Whatever one’s conclusions about the validity of MW’s statistical approach, we would argue that any meta-analysis claiming as its goal the attempted replication of an earlier meta-analysis on the same type of studies should use the same statistical test of significance, especially when that test of significance is the most accurate one available.

We acknowledge, however, that choice of statistical test does not change the precipitous drop in ES observed during the period that MW’s analysis covered. One possible explanation for this is offered by Bem et al. (BPB; 2001):

The z scores of the studies in the Milton-Wiseman database are significantly heterogeneous, and one of the observations made during the online debate was that several studies contributing negative z scores to the analysis had used procedures that deviated markedly from the standard ganzfeld protocol. Such a development is neither bad nor unexpected. Many psi researchers believe that the reliability of the basic procedure is sufficiently well established to warrant using it as a tool for the further exploration of psi. Thus, rather than continuing to conduct exact replications, they have been modifying the procedure and extending it into unknown territory. Not unexpectedly, such deviations from exact replication are at increased risk for failure. For example, rather than using visual stimuli, Willin modified the ganzfeld procedure to test whether senders could communicate musical targets to receivers. They could not. When such studies are thrown into an undifferentiated meta-analysis, the overall effect size is thereby reduced, and perversely, the ganzfeld procedure becomes a victim of its own success. (p. 208)

Rather than attempt to resolve what constituted “standard” ganzfeld procedure, BPB took the experimental route. They tasked three blind raters, each unfamiliar with the study outcomes, to rate the 40 studies in their database (from which all results had been erased), according to a 7-point standardness scale (where 7 indicated the greatest conformity to PRL protocols, and 1 indicated the least). As their guide to defining standardness, they were given the two original PRL ganzfeld papers of Bem and Honorton (1994) and Honorton et al. (1990).

However, Wiseman writes that BPB added a standardness measure that was not found in the papers given to the blind raters: participant selection. Measures such as “prior meditation experience,” “artistic or creative,” and “mental discipline practice,” Wiseman claims, were post hoc conditions based on knowledge of experimental outcomes, and therefore examples of retrospective data selection. But Wiseman is likely misled on this point. Bem and Honorton (1994) make it clear that replications should use participants of the selected type, probably because 100% of the PRL participants were selected in at least one of the previously mentioned ways, or others (e.g., strong belief in psi, friends, biologically related). Wiseman additionally went on to say, in a 2010 talk based on his article (Wiseman, 2010b), that the use of no sender was also treated as standard, even though the ganzfeld was considered a telepathy methodology. However, he did not mention that the PRL studies gave the participants the option of whether or not to have a sender, and four participants opted to have no sender. Honorton never claimed the ganzfeld method was only for testing telepathy; he would tell participants that it was for testing telepathy because it would seem more plausible to them than clairvoyance or precognition, therefore giving them more motivation. Donald McCarthy (1993) wrote of Honorton:

He told me, not long ago, that in designing the ganzfeld procedure, a primary reason for his choosing a telepathy protocol was that it might lead to more ready acceptance, since people seemed less threatened by the idea of “mental radio” than by other ways of conceptualizing psi. (p. 9)

The end result of the BPB analysis was strikingly in accordance with expectation; the success of the studies correlated significantly with the measures used to evaluate compliance with PRL protocols. Those studies that ranked above 4.0 on the scale (the midpoint) yielded significant results at a hit rate of 31.2% (1,278 trials, 29 studies, exact binomial \( p = .0002 \), one- tailed), and those that fell below gave a hit rate of 24% (n.s.). More dramatically, the studies that went to 6 or above (974 trials, 21 studies) scored a 33% hit rate (exact binomial \( p = 1.58 \times 10^{-8} \), or odds
against chance of 63 million to 1)—almost exactly PRL’s own. We argue that this is another example of a surprising consistency in psi research.

A possibility, however, that may have confounded the conclusions of the BPB meta-analysis, pertains to selected participants. Because Milton and Wiseman never did a heterogeneity analysis, they did not notice that their database was significantly heterogeneous by Timm’s chi square test on $z$ scores, and significantly heterogeneous by Honorton’s chi square test on effect sizes ($p = .07$, two-tailed; the alpha for chi square tests on meta-analyses is $p \leq .05$ to compensate for low statistical power). This led one of us (Derakhshani) to test the hypothesis that the source of the heterogeneity might have been the difference in scoring rates between selected and unselected participants. Given that the PRL database used only selected participants, it could be argued that a meaningful indicator of replication would be a comparison between the hit rates of the PRL studies and the MW studies that used selected participants. In fact, for the 513 trials in MW’s database from studies that used selected participants, 157 hits were obtained, for an overall hit rate of 30.6% (exact binomial probability of $p = .002$, one tailed), which is not significantly different from the 32.2% hit rate of the PRL studies (Fisher’s exact $p = .65$, two tailed). By contrast, the 661 trials with unselected participants produced a 24.7% overall hit rate, which is significantly different from that of the selected participants (Fisher’s exact $p = .014$, one-tailed) and, incidentally, also nonsignificantly different from the 27.3% hit rate of unselected participants in Storm et al. (2010) with four-choice designs; $p = 0.40$. Thus, it can be argued that Milton and Wiseman (1999) actually replicated the PRL results, if one considers roughly homogeneous populations alone—without even factoring in the relevance of homogeneous procedures—although the greater accuracy obtained through the standardness ratings in BPB suggests that study procedures still play a role in moderating outcomes.

In any case, as Storm et al. (2010) and many other meta-analyses (Bem et al., 2001; Derakhshani, 2014; Radin, 2006; Storm & Ertel, 2001; Storm et al., 2010; Tressoldi, 2011; Uts, Norris, Suess, & Johnson, 2010) demonstrate, the overall hit rate of the post-PRL database remains highly significant after the MW meta-analysis. Dean Radin (2010) makes an important point about this—namely that the controversy over replication in the ganzfeld (and in other psi paradigms) has advanced beyond the replicability of individual studies and is now about the replicability of experiments considered in groups. Not only are there individual meta-analyses confirming (what we have argued to be) a reasonable rate of replication across a wide swath of experiments over periods of several years; there are now groups of meta-analyses confirming consistency over many thousands of trials, in more than a hundred studies, and over five decades.

The Declining Decline Effect

In the last section of his article, Wiseman (2010a) writes:

The alleged psi effects associated with a certain procedure frequently have a curious habit of fading over the course of repeated experimentation. Skeptics argue that this is due to the parapsychologists identifying and minimizing potential methodological and statistical flaws over time. However, some parapsychologists have come up with creative ways of explaining away this potential threat, arguing that such decline effects are either an inherent property of psi or that psychic ability really does exist but is inversely related to the level of experimental controls employed in a study. (pp. 37–38)

As in the example of experimenter expectancy, we believe that Wiseman has left out several important observations for the decline effect. Moreover, we have already pointed out that the hypothesis from the preceding quote that “psychic ability . . . is inversely related to the level of experimental controls employed in a study” is not supported by the evidence available on a meta-analytic level, for most of the kinds of effects examined by experimental psi research. Additionally, in tandem with experimenter expectancy, decline effects are far from unique to parapsychology.

For example, Jonathan Schooler (2011), professor of psychological and brain sciences at the University of Santa Barbara, covered a number of examples of the decline effect in a debate over psi at Harvard, showing that they occur in research on schizophrenia, with medicines such as Pravastatin, Timolol, and Latonoprost, and even ecological relationships. Journalist Jonah Lehrer (2010) also wrote about the decline effect in a controversial article published in the New Yorker, which discussed the phenomenon’s occurrences in tests of the drug Zyprexa,
Beyond the Coin Toss

psychological effects such as verbal overshadowing (Schooler’s own widely cited research), biological correlations between asymmetry and mutation, and other paradigms. In order to contextualize the decline effect, Schooler (2011) proposed the following framework during a debate about psi at Harvard:

Controversial Prediction

*Mainstream view clearly most parsimonious at present but uncertain until decline effect is adequately understood. Need a process for recording all negative and unpublished findings to resolve issue*

An additional hypothesis for the decline effect not touched upon by Schooler in the above framework is the possibility of increased publication bias around the time an effect is first reported to be produced—as mentioned by Lehrer (2010). Indeed, Harris and Rosenthal (1988) illustrate this: They predicted in their assessment of Honorton’s (1985) ganzfeld meta-analysis that, taking into account corrections for minute publication bias, along with corrections for statistical errors and reporting errors, the true ganzfeld hit rate would decrease from 38% to around one third. This prediction was strikingly confirmed by the PRL meta-analysis (Bem & Honorton, 1994), which found a 32% overall hit rate across its 10 studies.

So the skeptical hypothesis mentioned by Wiseman for the decline across time is likely close to the truth for the drop in effect size from the earliest ganzfeld database to the second. But does it account for the decline in effect size from the PRL to the Milton and Wiseman (1999) database? Although declines can come about for a number of reasons, as regards the MW database two possible explanations are (a) treatment of exploratory research as confirmatory and (b) a change in the population tested over time. There is strongly supportive evidence for both. However, there is no available evidence that this decline can be explained by higher-quality research in the MW database, relative to the PRL database. This explanation also fails to account for the subsequent incline from the MW database to the post-MW database, or the significant incline across both, if they are considered together: $r = .27, p = .03$.

What about just the ganzfeld studies after the MW meta-analysis? To address this, we looked at data from the most recent post-MW ganzfeld database of 30 studies, from 1997–2008. We examined study ES versus study year and study ES versus study quality ratings, both for the entire database of 30 studies and for the two homogeneous subgroups of selected and unselected participants identified by Derakhshani (2014).

In plotting study effect size vs. study year, we found no decline in effect size ($r = 0$). Plotting study ES vs. study quality ratings, we found a positive and significant correlation, $r(28) = .37, p = .045$, two-tailed. That is, studies rated as having higher methodological quality produced larger effect sizes than lower quality studies, and this trend was statistically significant. For the entire database, then, the evidence appears to contradict Wiseman’s hypothesis.

Of course, one might reasonably argue that if the entire database is heterogeneous in ES distribution, then the correlations involving effect size may be misleading. Derakhshani (2014) indeed found highly significant
heterogeneity \((p = .002, \text{two-tailed})\) via the chi square test. He also found that blocking the studies according to
whether they used selected or unselected participants produced two safely homogeneous databases. The selected
participant subgroup consisted of 14 studies of four-choice design, with a 40.1\% overall hit rate across 748 trials;
and the unselected participant subgroup consisted of 15 studies of four-choice design (Roe & Flint, 2007, was
excluded because it used an eight-choice design) with a 27.3\% overall hit rate across 886 trials. Furthermore, the
difference in hit rates was extremely significant, Fisher’s exact \(p < .0001\).

A reasonable explanation for this is that selected studies have lower average quality ratings, but as Der-
akshani (2014) shows, this is not the case. Selected participants in fact produced a sample size weighted mean
quality rating of \(q = .84\) (where 1 is the highest possible rating), whereas the unselected participant studies produced
a lower mean quality rating \(q = .79\).

We did find that there was a small negative correlation between ES and study year for the selected partic-
ipant studies, but it was not significant, \(r(12) = .30, p = .29, \text{two-tailed}\). Moreover, we found a positive, nonsignif-
icant correlation, \(r(12) = .27, p = .37, \text{two-tailed}\), between study quality and study ES. We also found a positive
and nonsignificant correlation, \(r(12) = .26, p = .37, \text{two-tailed}\), between study quality rating and study year. Our
analyses thus do not support a relationship between quality and ES. More selected studies will be needed before we
can ascertain whether the positive correlation between these variables is real or spurious.

The unselected participant subgroup, on the other hand, had more striking results. For ES vs. year, we found
a highly significant positive correlation, \(r(14) = .65, p = .007, \text{two-tailed}\). For quality ratings vs. year, we found an
extremely significant positive correlation, \(r(14) = .86, p < .00002, \text{two-tailed}\). And for quality ratings vs. ES, we
found a large but nonsignificant positive correlation, \(r(14) = .40, p = .13, \text{two-tailed}\).

On the basis of these results, we can say with confidence that there is no decline in ES for unselected stud-
ies, and that there is no evidence for one in selected participant studies. We admit that we do not know why the find-
ings are so robust for the unselected participant subgroup, and nonsignificant for the selected participants subgroup,
but the question surely merits further research.

Given our reliance on the quality criteria of Storm et al. (2010), however, a skeptical reader might reason-
ably ask if there could be something problematic with or implausible about how these were constructed/judged.
Storm et al.’s (2010) quality ratings were made by two judges (graduate students of Tressoldi) who saw only the
method section of each article they assessed; all identifiers had been deleted, such as article title, authors’ hypothe-
ses, and references to results of other experiments. The seven criteria by which they evaluated the quality of a study
are reproduced for convenience below (Storm et al., 2010, p. 474):

1. Appropriate randomization (using electronic apparatuses or random tables).
2. Random target positioning during judgment (i.e., target was randomly placed in the presentation with
decoys).
3. Blind response transcription or impossibility to know the target in advance.
4. Number of trials pre-planned.
5. Sensory shielding from sender (agent) and receiver (perceiver).
6. Target independently checked by a second judge.
7. Experimenters blind to target identity.

Two judges answered “yes” or “no” to each of the criteria. Study quality was defined as the ratio of points
awarded with respect to the items applicable (minimum rating was \(1/7 = 0.14\); maximum rating was \(7/7 = 1\)), and
the quality ratings of each judge were averaged together. Storm et al. (2010) reported a Cronbach alpha for the
two judges’ ratings of .79, indicating high interrater reliability. Their criteria for study quality and their method
of determining quality scores seem reasonable to us, and we can see no major flaws in them that might nullify our
conclusions, for either the unselected or selected participants subgroup.
Beyond the Coin Toss

To summarize, the selected participants subgroup shows no evidence for the skeptical hypothesis, and the unselected participant subgroup shows some evidence against. As we can find nothing evidently wrong with the criteria and methods by which Storm et al. (2010) determined study quality, we conclude that the skeptical hypothesis suggested by Wiseman is inconsistent with the data, at least when it comes to the ganzfeld paradigm after the PRL experiments.

Beyond the ganzfeld, the forced-choice ESP meta-analysis by Storm et al. (2012) found a positive and highly significant incline effect for study year vs. study ES in their homogeneous database of 72 studies from 1987–2010, \( r = .31, p = .007 \), two-tailed, along with a positive and significant correlation between study year and study quality rating in their heterogeneous 91-study database, \( r = .25, p = .02 \), two-tailed. They also found a very weak, negative, and nonsignificant correlation between quality rating and ES, \( r = -0.08, p = .45 \), two-tailed. Honorton and Ferrari’s (1989) assessment of the forced-choice precognition literature, likewise, found that ESs had remained relatively constant through 1936–1987, although quality had substantially improved.

In sum, we find little evidence in either ganzfeld or forced-choice experiments for problematic decline effects. There is significant evidence in recent ganzfeld work for an incline across the MW database to the post-MW database, exactly no incline among all post-MW studies, and highly significant evidence for an incline just within unselected subjects studies for the post-MW database. As we have noted, interesting questions remain to be pursued, such as (a) what are the true correlations for ES vs. quality ratings, ES vs. year, and quality ratings vs. year, in the ganzfeld selected participant subgroup, and (b) why are the correlations so strongly significant for the unselected participants subgroup but nonsignificant for the selected participants subgroup? Such research questions, however, were not present in Wiseman’s discussion of declines.

The Progress of Parapsychology

Below, we present Wiseman’s perspective on the history of psi research and contrast it with our own. Wiseman (2010a) states:

Initial work, conducted between the early 1930s and late 1950s, primarily involved card guessing experiments in which people were asked to guess the identity of specially printed playing cards carrying one of five simple symbols. By the mid-1960s parapsychologists had realized that such studies were problematic to replicate and so turned their attention to dream telepathy and the possibility of participants predicting the outcome of targets selected by machines. In the mid 1970s and early 1980s, the ganzfeld experiments and remote viewing took over as dominant paradigms. In 1987, a major review of the area by parapsychologists K. Ramakrishna Rao and John Palmer argued that two sets of ESP studies provided the best evidence for the replicability of psi: the ganzfeld experiments and the differential ESP effect (wherein participants apparently score above chance in one condition of an experiment and below chance in another). More recently, parapsychologists have shifted their attention to alleged presentiment effects, wherein participants appear to be responding to stimuli before they are presented. Finally, there are now signs that the next new procedure is likely to adopt a neuropsychological perspective, focusing on EEG measurements or functional MRI scans as people complete psi tasks. (p. 39)

According to our assessment of the literature, the reason for the shift in research focus from (for example) dream ESP to the ganzfeld, and from the ganzfeld to presentiment, is the goal of finding an experimental paradigm that produces the largest ES for the least financial cost and time per trial. The ganzfeld, for example, is well known to produce comparable ESs to dream ESP but for a fraction of the time and cost per session (a typical ganzfeld trial typically takes 1–2 hours, compared to a full 24 hours for a dream ESP trial), and presentiment produces comparably greater effect sizes to ganzfeld on average; random effects ES = 0.13 (Tressoldi, 2011) and random effects ES = 0.21 (Mossbridge et al., 2012), respectively—but for even less time and cost per trial (typically a few seconds or minutes compared to 1–2 hours).

Even so, we emphasize that each experimental approach has its advantages—both for producing psi and understanding it—and that research in other paradigms certainly has not ended. The recent meta-analysis of forced-choice ESP studies by Storm et al. (2012) shows, for example, that 91 studies of admissible methodological quality were conducted from 1987–2010. By comparison, for the ganzfeld, only 60 such studies were found in the Storm et
al. (2010) meta-analysis, from 1987–2008. Despite the rise of the presentiment paradigm in the late 1990’s, moreover, Storm et al. still reported 30 ganzfeld studies conducted in the period from 1997–2008, exactly matching the number of studies in the Milton-Wiseman (1999) database from before the rise (1987–1997). These observations suggest that some research paradigms have been only minimally affected by the emergence of others.

In sum, our analysis leads us to reject both of Wiseman’s claims about psi research: that (a) decline effects consistently lead to nonreproducible results, and (b) parapsychologists routinely abandon old experimental procedures for new ones. On the contrary, we found evidence that research continues in the majority of parapsychology paradigms, and we think that there are more persuasive reasons than replication failure for why parapsychologists work to pioneer (and adopt) new research techniques.

A Couple Suggestions for the Way Forward

In Wiseman’s (2010a) concluding section, he writes:

To help the field move forward and rapidly reach closure on the psi question, parapsychologists need to make four important changes in the way they view null findings. First, they should stop trying lots of new procedures and cherry-picking those that seem to work and instead identify one or two that have already yielded the most promising results. Second, rather than varying procedures that appear successful, they should instead have a series of labs carry out strict replications that are both methodologically sound and incorporate the most psi-conducive conditions possible. Third, researchers should avoid the temptation for retrospective meta-analysis by pre-registering the key details involved in each of the studies. And finally, researchers need to stop jumping ship from one experimental procedure to another and instead have the courage to accept the null hypothesis if the selected front-runners don’t produce evidence of a significant and replicable effect. (p. 39)

Although we hope to have shown that the charges of “cherry-picking” new procedures—on the basis of the examples examined—are questionable, we do agree with Wiseman that parapsychology could benefit from focusing its resources on fewer research paradigms and using the best meta-analytic data on these paradigms to boost ESs and replication rates as high as possible. Here, we would like to make our own humble suggestions for how this could be done, according to our review of the evidence.

First, Derakhshani (2014) advises, on the basis of a predictive power model utilizing existing meta-analyses of ganzfeld studies, that it should be possible to boost the replication rates of future ganzfeld studies from ~30% to as high as 80%, while keeping the mean sample sizes of ganzfeld studies effectively the same, by the careful and exclusive use of selected participants in all (or as many possible) future ganzfeld studies. We want to be explicit: For the most recent Storm et al. (2010) database, the selected participant hit rate of 40% suggests that the recipe exists for a ganzfeld study with a significantly amplified effect size and chance of success. But care is necessary. This hit rate differs greatly from that of selected participants from previous databases (30% for the MW and 32% for the PRL) and therefore is likely not explicable by just any selection process.

In a similar vein, using Storm et al.’s (2010) selected participant hit rate of 40%, the required sample size is just 56. We emphasize that the characteristics of this selected sample have not been systematically reviewed; nevertheless, we can identify two highly powered studies in Storm et al. (2010) that provide a model for future investigators. Dalton (1997), using preselected artistic participants with positive attitudes towards psi and previous psi experiences, obtained a 47% hit rate in 128 trials (and also had the highest quality rating of 1.00 in Storm et al’s
between ratings was not significant (selected participant studies mean rating = 0.69, $q$ lower than the mean quality rating for studies using unselected participants ($\text{mean ES} = 0.03$. Derakhshani also found that the mean quality rating of the studies using selected participants was 0.73. By comparison, the 80 studies using unselected participants produced only 21/80 (26.3%) significant studies, with 11 studies using selected participants that produced a mean $\text{ES} = 0.09$, with 8/11 (73%) significant at the .05 level. Furthermore, for the more recent Storm et al. (2012) forced-choice meta-analysis, one of us (Derakhshani) found that at least 80% of such studies should reach significance at the .05 level.

If parapsychologists keep their studies relatively standard methodologically (given the positive correlation between standardness ratings and effect sizes in Bem et al., 2001), of high quality (given the positive correlation between quality and effect size found by Derakhshani), and use only well-selected participants, we predict the possibility of replication rates of 80% or greater. Large sample sizes also seem advisable given the positive correlation between $N$ and $\text{ES}$ found by Derakhshani for selected participants studies. In our view, such success would go a long way towards persuading the mainstream to replicate these studies.

Second, with regard to nonganzfeld experiments, in the Honorton and Ferrari (1989) forced-choice precognition meta-analysis, a subset of “optimal studies” in the homogenous database using selected subjects and trial-by-trial feedback had remarkably high replication rates—of the eight optimal studies, seven (87.5%) produced significant outcomes at the .05 level, with mean $z = 6.14$ and mean $\text{ES} = 0.06$. In the heterogeneous version of their database, there were 17 optimal studies, 15 (88%) of which reached statistical significance at the .05 level, with combined $z = 15.84$ and mean $\text{ES} = 0.12$—about twice the ES of ganzfeld studies using unselected participants. By comparison, the nine sub-optimal studies in their homogeneous database produced no significant results; mean $z = -1.29$ and mean $\text{ES} = 0.005$; and the optimal studies had significantly higher quality ratings than the sub-optimal studies (optimal mean $= 6.63$, $SD = 0.92$; suboptimal mean $= 3.44$, $SD = 0.53$; $t(10) = 8.63$, $p = 3.3 \times 10^{-4}$ two-tailed). Furthermore, for the more recent Storm et al. (2012) forced-choice meta-analysis, one of us (Derakhshani) found 11 studies using selected participants that produced a mean $\text{ES} = 0.09$, with 8/11 (73%) significant at the .05 level.

By comparison, the 80 studies using unselected participants produced only 21/80 (26.3%) significant studies, with mean $\text{ES} = 0.03$. Derakhshani also found that the mean quality rating of the studies using selected participants was lower than the mean quality rating for studies using unselected participants ($q = 0.69$ vs. $q = 0.81$), but the difference between ratings was not significant (selected participant studies mean rating $= 0.69$, $SD = 0.23$; unselected participant studies rating $= 0.80$, $SD = 0.21$; $t(89) = 1.61$, $p = .11$, two-tailed). In addition, 6 of the 11 selected participants studies produced a mean quality rating of 0.86, with no study rated less than 0.80, and yet the mean $\text{ES}$ of 0.08 for these six studies was still more than triple the mean effect size of the unselected participants studies. In addition, Derakhshani noticed that the 11 selected participants studies produced a strong correlation between $N$ and $z$; $r = .64$, $p = .012$, one-tailed, as would be expected under the assumptions of power analysis. By comparison, the 80 unselected participant studies produced a null correlation ($r = 0$). On the basis of these findings, we suggest that now is a suitable time for parapsychologists to reinvest in the forced-choice paradigm, prospectively plan a large set of optimally designed forced-choice ESP studies, and set the sample sizes for these studies large enough that they can expect at least 80% to reach statistical significance, assuming one of the observed effect sizes for optimal studies (we recommend using the conservative lower effect size estimate of 0.055). The expected outcome would then be that at least 80% of such studies should reach significance at the .05 level.

We also suggest that experimenters make every use of venues such as the Koestler Parapsychology Unit Registry (http://www.koestler-parapsychology.psy.ed.ac.uk/TrialRegistry.html) to preregister experiments, in order to circumvent what little publication bias may still exist in parapsychology. Additionally, we recommend careful examination of all methodological and statistical guidelines suggested by the program chairs of the Parapsychological Association 56th annual convention, originally put forward by Utts and Tressoldi (2013). Finally, we strongly recommend that any future large prospective studies make use of the two safeguards against experimenter misconduct proposed by Kennedy (2014): registering a multiple-experimenter protocol with independent copies of study outcomes, so as to prevent tampering; and providing the raw data for analysis by others after a study is completed.

If parapsychologists were to adopt these suggestions for ganzfeld and forced-choice ESP studies and produce the outcomes predicted on the basis of these meta-analytic findings, we believe it would go a long way towards convincing the mainstream academic community to take seriously the scientific possibility of ESP, as well as to invest resources to attempt large-scale replications of the results. Conversely, if the predicted outcomes were grossly disconfirmed, it would raise serious doubts about the positive results of past experiments. Either outcome, in our view, would constitute significant progress in the scientific assessment of whether or not ESP exists.
Final Thoughts

Replying to a series of critiques of psi research similar to Wiseman’s, Honorton (1993) wrote that counter-advocates of parapsychology had drifted away from making active contributions to the field—some of which in the past (e.g. Hyman & Honorton, 1986) were substantial:

Critics have been forced to admit that parapsychology has demonstrated anomalous effects that need to be explained and they have run out of plausible conventional explanations . . . instead, they offer a caricature of the history of parapsychology and present polemical arguments designed to convince us that there is really nothing in parapsychology that warrants scientific interest, except, perhaps, for the motivations of those who persist in studying it. (Honorton, 1993, p.191)

In reviewing Wiseman’s essay, Heads I Win, Tails You Lose: How Parapsychologists Nullify Null Results, we have come to the opinion that it too presents such a caricature. Wiseman’s portrait of parapsychology is simply more dismal than the data. The impression one is left with after reading it, not uncommon in published criticisms of psi research, is that parapsychology is an unprofitable area of inquiry, from which little is to be gained but frustration with the caprice of psi and the confirmation bias of psi researchers. This vignette is regrettable in our opinion, and poorly serves both parapsychology and organized skepticism; for whereas the former is depleted of its financial and human support as a result of negative publicity, the latter is—just as importantly—deprived of a unique opportunity to examine the evidence for psi, for the same reason.

We hope to have shown that the opportunity is still there. The evidence for some forms of psi is stronger than we would expect it to be; it is not easily dismissed, easily ignored, or, indeed, easily summarized. Our approach considered and responded to each general criticism raised by Wiseman with specific evidence from the literature, and showed that parapsychology has reason to intrigue all participants in the psi debate, from advocates to counteradvocates to (like us) newcomers becoming familiar with the field. On the other hand, by putting forth recommendations, we have tried to demonstrate that parapsychology can still improve its face validity. High-powered prospective studies, sizeable proportions of significant results, large effect sizes, open source dissemination of data, pre-registered studies, multiple-experimenter protocols, and much more, are perhaps just around the corner.

It is this commitment to improve, in fact, where both sides can meet, for we believe the joint goal must be to produce the highest quality parapsychology database within our present means—a database that we have argued is strongly suggested by the research. If we maintain a willingness to improve the evidence to this degree then perhaps the debate will advance “Beyond the Coin Toss”—dispensing with heads, tails, and the single outcomes of winning and losing, implied in the titles of Wiseman’s (2010a) essay and Carter’s (2010a) response, for a collaborative attempt to resolve the psi enigma for the twenty-first century.

References


Abstracts in Other Languages

German

JENSEITS DES MÜNZENWURFS: EINE NACHPRÜFUNG VON WISEMANS KRITIK AN DER PARAPSYCHOLOGIE


Spanish

MÁS ALLÁ DE LANZAR UNA MONEDA : UN EXAMEN DE LA CRÍTICA A LA PARAPSICOLOGÍA DE WISEMAN

RESUMEN: Examinamos la crítica de la parapsicología ofrecida por el Profesor Richard Wiseman en su artículo de 2010, Heads I Win, Tails You Lose; How Parapsychologists Nullify Null Results, publicada en Skeptical Inquirer, y ofrecemos refutaciones detalladas de sus principales argumentos. Algunos de los análisis que llevamos a cabo son: Comparamos la reproducibilidad de los experimentos psi con la reproducibilidad de los experimentos en campos convencionales semejantes, mostrando que son equivalentes. Utilizando enfoques tanto teóricos como empíricos demostramos que los efectos de archivo (filedrawer effect) no están presentes en el ganzfeld. Examinamos y criticamos los casos de supuesta anulación por el experimentador de resultados nulos. Las conclusiones del meta-análisis de Milton y Wiseman son criticadas en base a los resultados de Bem, Palmer, y Broughton, así como de nuestros propios resultados. Examinamos y rechazamos los efectos de disminución ostensible en los paradigmas ganzfeld y de elección forzada de ESP. Finalmente, presentamos estrategias de progreso de acuerdo con las tendencias más atractivas y consistentes que hemos encontrado en los datos contemporáneos. Presentamos un análisis de la crítica en la parapsicología, con Wiseman como el ejemplo principal, que muestra el grado en el que la literatura no apoya las aseveraciones escépticas, así como la forma en que parece apoyar a psi como la explicación más plausible de los datos.
AU-DELA DU PILE OU FACE : EXAMEN DES CRITIQUES DE LA PARAPSYCHOLOGIE PAR WISEMAN

RESUME : Nous examinons la critique de la parapsychologie proposée par le Professeur Richard Wiseman dans article de 2010, Heads I Win, Tails You Lose; How Parapsychologists Nullify Null Results, publié dans le Skeptical Inquirer, et nous proposons une réponse détaillée à ses principales critiques. Parmi les analyses que nous avons conduites se trouvent les suivantes : nous avons comparé la reproductibilité des expérimentations psi à la reproductibilité des expérimentations dans d’autres champs de recherche mainstream, en découvrant qu’elles étaient équivalentes. En utilisant des approches à la fois théorique et empirique, nous avons démontré que les effets de fonds de tiroir ne sont pas présents dans le ganzfeld. Des cas de supposée nullification de résultats nuls par des expérimentations sont examinés et critiqués. Les conclusions de la méta-analyse de Milton et Wiseman sont confrontées aux résultats de Bem, Palmer et Broughton, ainsi qu’à nos propres résultats. D’apparents effets de déclin dans les paradigmes du ganzfeld et de la perception extra-sensorielle à choix forcée sont testés et rejetés. Finalement, nous présentons des stratégies pour progresser en nous basant sur les tendances les plus intéressants et cohérents découvertes dans notre base de données. Nous présentons un examen de la critique de la parapsychologie en prenant Wiseman comme exemple, montrant à quel point la littérature scientifique échoue à soutenir les revendications des sceptiques et à quel point elle fait de l’hypothèse psi l’explication la plus plausible des données.
HOW TO REMOVE THE INFLUENCE OF EXPECTATION BIAS IN PRESENTIMENT AND SIMILAR EXPERIMENTS: A RECOMMENDED STRATEGY

BY JAN DALKVIST, JULIA MOSSBRIDGE,* AND JOAKIM WESTERLUND

ABSTRACT: Here we reconsider expectation bias in so-called presentiment experiments, with focus on how to handle it. In such experiments, presentiment is usually thought to be demonstrated by showing that significant physiological differences precede stimuli presumed to give rise to different arousal levels. Often these differences suggest that physiological arousal is more likely to precede arousing rather than calming stimuli. Conceivably, however, such reactions can be explained as resulting from expectation bias of the gambler’s fallacy type. This bias is based on the (false) notion that the likelihood of an arousing stimulus being presented grows as the number of consecutive calming stimuli increases. Different ways of controlling or avoiding the bias are discussed. Our resulting recommendation is to use analysis of variance (ANOVA) to separate the effect of the bias from the hypothetical presentiment effect, preferably at the trial-by-trial level. We also recommend applying ANOVA to each participant separately and using a “counting” method to test for possible presentiment effects at the group level. Application of ANOVA is illustrated using a simulated example. We anticipate ANOVA can handle not only the gambler’s fallacy bias but also similar biases, in presentiment experiments as well as in some conscious precognition experiments.

Keywords: expectation bias, gambler’s fallacy, presentiment, precognition, analysis of variance

Imagine that you are running an experiment at the casino with a woman named Mary playing roulette. Mary, like most people, is a victim of the “gambler’s fallacy,” and she believes that, for each time the ball has dropped into a red slot, the chance increases that the ball will drop into a black slot the next time. Imagine also that Mary has decided to bet on black 10 times in a row, and that for each successive time the ball has dropped into a red slot, her heart rate increases by, say, one beat per min (due to excitement) until the ball drops into a black slot, whereupon her heart rate returns to baseline. After the experiment, you compute the mean number of heart beats at the moment before the ball dropped into a red slot as well as the mean number of heart beats at the moment before the ball dropped into a black one.

What is the statistically expected difference between the two measures? If you ask a statistically naïve person who, like Mary, is not familiar with the gambler’s fallacy, the answer will be something like this: “The mean heart rate will be higher before the ball drops into a black slot than before it drops into a red one, because Mary’s heart rate will reach a peak at that moment.” But a statistically more sophisticated person, who is familiar with the gambler’s fallacy, may instead give you an answer like this one: “The mean number of heart beats before the ball falls into a black slot could in the long run by no means be different from the mean number of heart beats before the ball falls into a red slot. The reason? Each outcome (red or black slot) is independent of all other outcomes, and the two different outcomes have exactly the same likelihood of occurrence.”

Who is right? As two of us (JD and JW) have made clear before (e.g., Dalkvist & Westerlund, 2006), the surprising answer is that neither of them and both of them are right. The sophisticated respondent is right in that, in the long run, the means of the heart beats before the two alternative outcomes will be exactly the same—assuming, of course, that no presentiment, that is, conscious or unconscious psi-mediated predictions of future events, occur. However, the naïve respondent is also right: In the short run, when the length of a run is finite, the mean heart rate will be higher just before the ball falls into a black hole than just before it falls into a red one, even if no presentiment

1 This paper is a revision of a 2013 paper presented at the 56th Annual Convention of the Parapsychological Association, Viterbo, Italy.
occurs. However, the reason suggested by the naïve respondent is wrong: the reason is not the fact that arousal tends to reach a peak just before the ball drops into a black hole; it is instead related to inappropriate mean calculations, as will be shown below.

We will now leave the casino for the moment and turn to the parapsychological laboratory. Here we will take a look at a certain kind of precognition experiment that is conceptually identical to the above casino experiment. This type of experiment was originally invented by Dean Radin, and has come to be known as the “presentiment” experiment (Radin, 1997). In a typical presentiment experiment, instead of red and black slots, there are two types of pictures shown to the participants: emotionally activating and calming. The pictures are selected randomly, with replacement, from a pool. The participants’ physiological arousal is measured not only when a picture is presented, but also before the presentation. Sometimes heart rate has been used as the measure of arousal, but all sorts of physiological reactions reflecting emotional arousal have been tested—most often electrodermal activity. The hypothesis tested is that, as compared to a calming picture, for an emotionally activating picture, the arousal level becomes enhanced not only as the picture is being presented, but also 1–10 s before it is presented, thus exhibiting a sort of physiological “presentiment” of the imminent presentation of an emotionally activating picture. Although this hypothesis has received some support (for a review and meta-analysis, see Mossbridge, Tressoldi, & Utts, 2012), several alternative interpretations have been discussed (Dalkvist & Westerlund, 2006; Mossbridge, Tressoldi, & Utts, 2012; Radin, 2004), most frequently effects of the gambler’s fallacy, considered here.

It is apparent on reflection that in the long run, that is, in an infinitely long sequence of randomly ordered activating and calming stimuli, the average arousal level preceding calming stimuli will be equal to the average arousal level preceding activating stimuli—provided, of course, that no presentiment is at work. Remarkably, however, this equality does not hold for a finite, that is, real sequence of calming and activating pictures, as suggested above. For such sequences, assuming the presentiment experiment is performed using a person who is subject to the gambler’s fallacy, even when there is no presentiment, the expected mean arousal level preceding activating stimuli is greater than the expected mean arousal level preceding calming stimuli. As would be expected, though, the difference between these means decreases as the length of the sequence increases, as will be shown below. Thus, for participants in a presentiment experiment who are subject to the gambler’s fallacy, a sequence-length-dependent bias is to be expected, short sequences tending to give rise to larger biases than long ones.

That the expectation bias is real is a recent insight, dating back to the Parapsychological Association conference in Paris in 2002, even though the possibility of an expectation bias already had been discussed by Radin (1997) and by Ertel (1997). At that conference, incontestable proofs of the bias were presented in two independent, complementary papers, one by Dalkvist, Westerlund, and Bierman (2002) and the other by Wackermann (2002). The former paper foremost used simulated “paper-and-pencil” demonstrations (some of which will be reconsidered below) as proofs and also provided a conceptual explanation of why the bias appears in the first place and behaves as it does. Wackermann demonstrated the bias using formal mathematical analysis.

Before these papers were presented, most parapsychological researchers did not even contemplate the possibility that a bias of the present type might exist, or, if they did, they thought it was illusionary. This belief was supported by a proof advanced by Bierman (1999), purporting to show that “sequential patterns cannot explain anomalous anticipatory physiological behavior” (p. 1), which, however, turned out to be restricted to infinitely long sequences. However, based on inspection of his own data, Radin (1997) did believe there was a bias caused by expectation, but he rejected it as being too small to have any practical consequences. However, this comforting conclusion was questioned by Ertel (1997), who recommended that the presumed bias should be tested systematically, based on a suggested conditioning model. However, there was little clear understanding of why and when there was a bias in the first place, or a more detailed description of how it behaved.

Fortunately, the bias can be tamed: It can be reduced, altogether avoided, or its effects can be isolated by statistical means. One purpose of the present paper is to compare alternative ways of handling the bias and argue for choosing one of them: using statistical means to isolate the effect of the bias from a possible presentiment effect, thereby rendering it harmless. Another purpose is to demonstrate concretely, using simulated data, how this statistical correction method can be applied in practice. Finally, in view of its wide and important potential implications, we want to urge our colleagues and other people to give the present expectation bias more attention than it has received so far.
The Bias

Before discussing alternative ways of handling the bias, we will take a look at a simulated presentiment experiment where no measure has been taken to reduce or eliminate the bias. The experiment is very short: a stimulus sequence involves only two pictures, a calming picture (C) and an activating one (A). This gives four possible sequences:

\[
\begin{align*}
C & \quad C \\
C & \quad A \\
A & \quad C \\
A & \quad A
\end{align*}
\]

Now let us assume that for each sequence we have just one participant and that each participant believes in the gambler’s fallacy, and, as a consequence, the participant’s arousal level increases by one unit in response to a calming stimulus. This situation is depicted in the left-hand column of Table 1. The participant who is to watch two calming pictures starts (like everyone else) with an arousal level of zero. After the first calming picture, this participant believes that the chance increases that the next picture will be activating, and arousal level rises accordingly, by one unit. Thus, as far as this participant is concerned, the second stimulus is preceded by one arousal unit.

Table 1

Between-Groups Analysis of Simulated Presentiment Data

<table>
<thead>
<tr>
<th>Arousal-modelled sequences</th>
<th>Stimuli</th>
<th>Activating</th>
<th>Calming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(\text{Sum}(^1A))</td>
<td>(n_A)</td>
</tr>
<tr>
<td>(^0C/C)</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(^0C/A)</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(^0A/C)</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(^0A/A)</td>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.25</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. \(^0A/C\) = zero arousal preceding an activating or a calming stimulus; \(^1A/C\) = one arousal unit preceding an activating or a calming stimulus; \(\text{Sum}(^1A)\) = sum of arousal units preceding activating stimuli; \(n_A\) = number of activating stimuli; \(\text{Mean}(^1A)\) = mean of arousal units preceding activating stimuli; \(\text{Sum}(^1C)\) = sum of arousal units preceding calming stimuli; \(n_C\) = number of calming stimuli; \(\text{Mean}(^1C)\) = mean of arousal units preceding calming stimuli.

When we first calculate the mean arousal level before activating pictures and the mean arousal level before calming pictures for each single participant and then average each of the two resulting sets of means, we find that the overall mean arousal level before activating pictures (\(\approx 0.33\)) is larger than the overall mean arousal level before calming pictures (\(\approx 0.17\)). Thus, by calculating the grand means of the arousal levels before activating and calming pictures, we get a bias, which in this particular case is substantial.
How to Remove the Influence of Expectation Bias

The design of this little experiment may be categorized as within-participant, as the same participants were assigned to two different conditions. However, this analysis deviates from that in a typical within-participant design in that the bias was calculated not as the mean of the individual arousal differences, as is customary in a within-participant design to reduce sampling errors, but as the difference of the group means of the individual arousal levels for the two stimulus conditions, as in a between-groups design. This choice was made to permit all four sequences to be used—even the CC- and AA-sequences—despite the fact that they give rise to undefined arousal means due to zeros in the denominators.

As will be shown in a moment, this deviation from a standard within-participant design does not account for the bias. Neither is an ordinary within-participant analysis a better method to handle the bias. This is because even though the random effect of inter-participant noise becomes zero in within-participant designs, the expectation bias as such does not necessarily get any smaller. This is shown in Table 2, where data from Table 1 are reanalysed using a within-participant analysis. In this analysis, the CC-sequence as well as the AA-sequence (the first and the last sequence) have been dropped, as can be seen. The reason for doing this is, of course, that in a within-participant analysis, not only must undefined parameters (ratios with zeros in the denominator) be excluded, as was done in the between-groups analysis, but also the whole sequence consisting solely of C-stimuli and the whole sequence consisting solely of A-stimuli, so that each sequence will contain at least one A-stimulus and at least one C-stimulus. As can be seen, the present analysis leads to an even larger bias than that shown in Table 1. Note that when the data are exactly the same in the two methods, the values of the expectation bias will also be exactly the same, even though expectation bias can be more easily confirmed statistically using the within-participant analysis, due to the random effect of interparticipant noise being reduced to zero.

### Table 2

<table>
<thead>
<tr>
<th>Arousal-modelled sequences</th>
<th>Stimuli</th>
<th>Activating</th>
<th>Calming</th>
<th>Mean(‘A) - Mean(‘C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sum(‘A)</td>
<td>nA</td>
<td>Sum(‘C)</td>
</tr>
<tr>
<td>0C1A</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0A0C</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>1/2 (Bias)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. 0A/0C = zero arousal preceding an activating or a calming stimulus; ‘A = one arousal unit preceding an activating stimulus; Sum(‘A) = sum of arousal units preceding activating stimuli; nA = number of activating stimuli; Mean(‘A) = mean of arousal units preceding activating stimuli; Sum(‘C) = sum of arousal units preceding calming stimuli; nC = number of calming stimuli; Mean(‘C) = mean of arousal units preceding calming stimuli.

Basically, the bias is due to a crucial difference between an arousal increment before a calming stimulus and one before an activating stimulus. The former arousal increment is preceded and followed by the same type of stimulus, a calming one, whereas an arousal increment before an activating stimulus is preceded by a stimulus of the opposite type, that is, a calming one. As a consequence of this difference, arousal increments before calming stimuli tend to cumulate in sequences with relatively many calming stimuli, whereas no corresponding cumulations of arousal increments occur before activating stimuli. This is illustrated in Table 3, where the eight possible sequences containing three calming or activating stimuli have been split into two groups: one dominated by calming stimuli (the left-hand group) and one dominated by activating stimuli (the right-hand group). Like Mary’s heart rate in the above casino experiment, arousal increases by one unit each time a calming stimulus appears and returns to baseline when a calming stimulus has been followed by an activating stimulus. As can be seen, all arousal increments...
before calming stimuli can be found in the group of sequences dominated by calming stimuli, whereas the arousal increments before activating stimuli are evenly distributed among sequences in the two groups. Such a difference between arousal increments before calming stimuli and activating stimuli implies that when arousal means are calculated for each sequence (or each subset of two or more sequences) separately, arousal increments before calming stimuli receive smaller weights than do arousal increments before activating stimuli, due to the occurrence of a larger number of stimuli in the denominator.

Table 3

<table>
<thead>
<tr>
<th>Dominance of calming stimuli</th>
<th>Dominance of activating stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^0C^1C^2C)</td>
<td>(^0A^0C^1A)</td>
</tr>
<tr>
<td>(^0A^0C^1C)</td>
<td>(^0A^0A^0C)</td>
</tr>
<tr>
<td>(^0C^1C^2A)</td>
<td>(^0C^1A^0A)</td>
</tr>
<tr>
<td>(^0C^1A^0C)</td>
<td>(^0A^0A^0A)</td>
</tr>
</tbody>
</table>

Note. C = Calming stimulus. A = Activating stimulus. Superscripts indicate arousal level. Arousal increases by one unit in response to a calming stimulus and returns to baseline, which is equal to zero, when an activating stimulus emerges.

But as the sequence length increases, by virtue of the law of large numbers, the possible sequences become more and more similar to each other in terms of relative number of calming and activating stimuli, as can be demonstrated by simulation, using, for instance, the standard deviation as a measure of dissimilarity among sequences. As a consequence of this increasing similarity among sequences, the bias decreases as sequence length increases, as demonstrated in Table 4. As can be seen in this table, both the mean of the average individual arousal level before activating stimuli and the mean of the average individual arousal level before calming stimuli increase as the sequence length increases, but at a higher level for activating stimuli than for calming ones. As can also be seen, the two means converge on the same value: 0.50—the expected value in the long run, meaning that the bias will decrease as the sequence length increases. (If the two sequences containing only activating or only calming stimuli are dropped, the mean for activating stimuli becomes 0.50 throughout.)

The above explanation of the expectation bias and its convergence on zero in the long run is based on the gambler’s fallacy, and is therefore, strictly speaking, limited to this particular type of bias. It should be a straightforward matter to extend the current explanation to include other types of biases that, like the gambler’s fallacy bias, are caused by the occurrence of different distributions of arousal changes (or lack thereof) for different types of stimuli.

Different Ways of Handling Expectation Bias

Six alternative methods for handling the expectation bias are listed and characterized in Table 5. The first method is perhaps the most common one. Using this method, one tests if there is any statistically significant bias, using, for example, regression analysis, and if no significant bias is found, the conclusion is drawn that there is no bias that one needs to take care of. This approach is not satisfactory, however, for two similar reasons. First, whether or not a statistically significant bias is found is not only dependent on whether there really is a bias, but also on the sample size and the strength of the bias, meaning that the method is uncertain as well as arbitrary. Second, even if no statistically significant bias is found, but only a tendency for a bias to occur, one does not know if, and how much, this tendency will affect a subsequent statistical test of the hypothetical presentiment effect. Thus, the current method is seriously affected by uncertainty.
The next three methods listed in Table 5 all have the same goal: to avoid the bias altogether. All of them have the drawback of requiring a large amount of participants or trials per participant, however.

### Table 4

**Simulated Average Individual Mean Arousal Levels for Activating and Calming Stimuli and Corresponding Biases for Sequences Ranging From 2 to 30 Trials Using a Binary Model and a Between-Groups Analysis**

<table>
<thead>
<tr>
<th>Sequence length (trials)</th>
<th>Stimuli</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activating</td>
<td>Calming</td>
<td>Bias</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.333</td>
<td>.168</td>
<td>.165</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.429</td>
<td>.238</td>
<td>.191</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.467</td>
<td>.283</td>
<td>.184</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.484</td>
<td>.316</td>
<td>.168</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.492</td>
<td>.343</td>
<td>.149</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.496</td>
<td>.361</td>
<td>.135</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.498</td>
<td>.377</td>
<td>.121</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.499</td>
<td>.390</td>
<td>.109</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.500</td>
<td>.401</td>
<td>.099</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.500</td>
<td>.409</td>
<td>.091</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.500</td>
<td>.417</td>
<td>.083</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>.500</td>
<td>.466</td>
<td>.034</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>.500</td>
<td>.467</td>
<td>.033</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Arousal increases by one unit in response to a calming stimulus and stays at that level until an activating stimulus appears.

The most obvious way of avoiding the bias altogether is simply to present each participant with only one stimulus (Method 2). But this requires a very large number of participants to bring the interindividual variation to a sufficiently low level to allow any existing presentiment effect to be confirmed statistically.

Another way of avoiding the bias altogether is to refrain from calculating arousal means across calming and activating stimuli, respectively, and just calculate and compare the corresponding sums (Method 3), which are not affected by any expectation bias at all (as can be seen from Table 1, where both arousal sums equal unity). In practice, this method can be applied either at the group level or at the individual level using an ordinary $t$ test to test for a difference between mean arousal sums before calming and activating stimuli, respectively. There are two weaknesses inherent in the just-sum method currently considered, however. One has to do with sampling of stimuli: because stimulus sampling is done with replacement, as is obligatory in any precognition experiment, an unequal number of calming and activating stimuli is to be expected in the sample due to chance (and in some experiments, by design), which, in turn, will give rise to corresponding unequal arousal sums before calming and activating stimuli. Also, a large variation caused by the previous stimulus is to be expected. At best, these weaknesses may be compensated for by using a large number of participants or a large number of trials to reduce the measurement errors caused by the two weaknesses.

Still another way of avoiding the bias altogether is to use sums (or, equivalently, means for pooled sequences) in a complete and balanced design (Method 4). That the design is complete means that all possible sequences of a given length occur, and that the design is balanced means that each sequence occurs equally often. The current method is illustrated in Table 6. Let us start by looking at the (optional) mean calculations. When calculating the mean arousal level before activating pictures and the mean arousal level before calming ones in this
case, we do not care about from which participant the data are obtained. Thus, we simply sum all the arousal values before activating pictures (= 1) and divide the resulting sum by the total number of activating pictures (= 4). The corresponding mean arousal level is one fourth. We do the same computations for the arousal values before calming pictures. In the present example, we get exactly the same mean as that obtained for the activating pictures. Thus, the difference between the two means is zero, and we do not get any bias at all.

Table 5

<table>
<thead>
<tr>
<th>Method</th>
<th>Goal(s)</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test if there is a statistically significant bias</td>
<td>To find out if there is a bias that needs to be taken care of</td>
<td>Uncertainty; arbitrariness</td>
</tr>
<tr>
<td>Use different participants in different single trials</td>
<td>To avoid the bias</td>
<td>Requires many participants</td>
</tr>
<tr>
<td>Use sums instead of means</td>
<td>To avoid the bias</td>
<td>Requires many participants or individual trials</td>
</tr>
<tr>
<td>Use sums (or means for pooled sequences) in a complete and balanced design</td>
<td>To avoid the bias</td>
<td>Requires many participants or individual trials</td>
</tr>
<tr>
<td>Use means for pooled sequences in an incomplete and unbalanced design</td>
<td>To reduce the bias</td>
<td>Uncertainty; may violate the assumption of independent responses</td>
</tr>
<tr>
<td>Use a two-way ANOVA</td>
<td>(a) To isolate the effect of a bias from the effect of a possible presentiment effect, thereby rendering it harmless; (b) to test for the existence of a presentiment effect independently of any expectation effect; (c) to neutralize the direct effect of the previous stimulus</td>
<td>May violate the statistical assumption of independent responses</td>
</tr>
</tbody>
</table>

But how can it be that the bias fails to appear in a complete and balanced design, despite the fact that the picture sequences are finite? The reason is simple: When the experimental design is complete and balanced and the data are pooled across sequences before averaging (i.e., the mean is calculated for all the individual’s arousal sums for calming and activating stimuli, respectively) the number of positive and negative pictures is exactly the same. In other words, comparing arousal means before calming and before activating pictures is equivalent to comparing the corresponding sums of arousal, which, as pointed out before, do not produce any bias. This means that when the design is complete and balanced and the data are pooled across sequences, calculating grand means (1/4 in Table 6) is not necessary; it is enough to calculate the grand sums (“1” in Table 6), as neither the grand means nor the grand sums are expected to differ between calming and activating stimuli. Thus, when the mean calculations are omitted, Method 4 is equivalent to Method 3, except that no errors due to stimulus sampling are to be expected in Method 4. (The design in Table 1, which does give rise to a bias, is also complete and balanced, but in that case arousal means are calculated for each sequence separately before the grand means are calculated.)

Unfortunately, for practical reasons, a complete and balanced design can generally not be obtained, due to the large number of participants that are required unless the sequence length is very small. Thus, as can be seen
from Table 7, if there is an equal probability for calming and activating pictures, we need at least 4 participants for a sequence length of 2 trials, 8 participants for a sequence length of 3 trials, 16 participants for a sequence length of 4 trials, and so forth. Soon, however, the number of participants will be prohibitively large, as can be seen on the bottom lines in Table 7.

Table 6
Analysis of Simulated Presentiment Data Using Complete and Balanced Pooling

<table>
<thead>
<tr>
<th>Arousal-modelled sequences</th>
<th>Stimuli</th>
<th>Activating</th>
<th>Calming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\sum(1A)$</td>
<td>$n_A$</td>
</tr>
<tr>
<td>$^0C' C$</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$^0C' A$</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$^0A' C$</td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$^0A' A$</td>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>Bias = $1/4 - 1/4 = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $^0A' /C = zero arousal preceding an activating or calming stimulus; $1A' /C = one arousal unit preceding an activating or a calming stimulus; $\sum(1A) = sum of arousal units preceding activating stimuli; $n_A = number of activating stimuli; $\sum(1C) = sum of arousal units preceding calming stimuli; $n_C = number of calming stimuli.

This fact brings us to the next method (Method 5) listed in Table 5: using means for pooled sequences in an incomplete and unbalanced design. This can be seen as a substitute for Method 4 when the required number of participants is too large for that method to be used. But even though Method 5 serves only to reduce the bias, not get rid of it, it may attenuate the bias substantially. It should be noted that Method 5 differs from Method 4 in using means instead of sums, thus avoiding the measurement errors that can be expected in any just-sum method unless the design is complete and balanced.

Although Method 5 may reduce the bias substantially, this method is not quite satisfactory, for at least two reasons. First, we do not know exactly how big the bias is, which, in turn, makes it impossible to estimate the size of the presentiment effect. Secondly, if one chooses Method 5, when using conventional statistical methods (in contrast to statistical simulations) to test for the existence of a presentiment effect, one runs the risk of violating the assumption of independent observations, as the within-participant responses (but not the between-participant responses) are statistically dependent.

The sixth and final method listed in Table 5 involves a two-way ANOVA, with the previous and the forthcoming stimuli as independent variables and the physiological measure as the dependent one. The major goal of using a two-way ANOVA is two-fold: (a) to isolate a possible expectation bias effect from a possible presentiment effect, thereby rendering the expectation bias harmless, and (b) to test for the existence of a presentiment effect, independently of any expectation bias.

As will be argued below, this method in general should be applied only to data from a single individual. If this is done, however, the method has the same problem as the pooling method just discussed (Method 4): When using standard statistical methods to test for the existence of a presentiment effect, one runs the risk of violating the assumption of independent observations, as the within-participant responses can be expected to be statistically dependent. However, particularly in view of the fact that methods exist to handle this problem (which we will consider in the Discussion), it may be outweighed by a unique advantage of the present method: It does not only separate any
expectation effect from a possible presentiment effect, but it also separates the direct effect of the previous stimulus from a possible presentiment effect, thus increasing the power to detect any existing presentiment effect.

Table 7
Minimum Number of Participants Required for the Bias to Disappear When All Data Are Pooled Across Participants

<table>
<thead>
<tr>
<th>Sequence length (Number of trials for each participant)</th>
<th>Number of participants required ( (2^n, \text{where } n = \text{number of trials for each participant}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>128</td>
</tr>
<tr>
<td>8</td>
<td>256</td>
</tr>
<tr>
<td>9</td>
<td>512</td>
</tr>
<tr>
<td>10</td>
<td>1024</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>20</td>
<td>(1049 \times 10^6)</td>
</tr>
<tr>
<td>30</td>
<td>(1074 \times 10^9)</td>
</tr>
<tr>
<td>40</td>
<td>(1100 \times 10^{11})</td>
</tr>
</tbody>
</table>

A Simulated ANOVA

To demonstrate how presentiment data can be handled using ANOVA, we have constructed a simple, idealized experiment involving only one participant, who is presented with a sequence of 12 activating (A) or calming (C) pictures. As far as the expectation effect is concerned, we have let arousal change in the same way as in the casino experiment with Mary, as can be seen here:

\[C^1, C^2, C^3, A^0, C^1, A^0, A^0, C^1, C^2, C^3, A^0, A^0\]

Thus, for each new calming stimulus that the participant encounters, we add one arousal unit until an activating stimulus appears, whereupon the arousal level drops to baseline, which, as before, is set to zero. (There are several alternatives to such a linear expectation/arousal model. One is the above-mentioned binary model (Table 4), which can be seen as a simplified version of the linear model, where arousal increases by a given amount in response to a calming stimulus and remains at that level until an activating stimulus brings arousal back to baseline. The most realistic model is probably a sigmoid model, meaning that arousal grows only slowly, or not at all, at the beginning of a series of calming stimuli and then increasingly fast as the number of calming stimuli increases, up to some inflexion point at which the curve starts to level off. However, when real data are tested using ANOVA, no specific arousal model needs to be assumed, as ANOVA is constructed to separate variance coming from different sources: It does not matter, for example, whether one of the two types of preceding stimuli acts to increase or decrease the arousal level as compared to the other type; both arousal differences will be separated from the presentiment effect by ANOVA. Thus, a model describing a changing arousal level as a function of expectation does not even need to conform to the gambler’s fallacy (an alternative conceivable bias will be considered in Discussion).
How to Remove the Influence of Expectation Bias

We have, of course, also entered a presentiment model into our simulated experiment. According to this model (not entered into the above sequence of \(A\)- and \(C\)-stimuli, which only shows the expectation effects), the participant has infallible presentiment: each time the next stimulus is activating, one arousal unit is added, and when the next stimulus is calming, the arousal level remains unaltered. As in the case of the expectation effect, however, where real data are tested using ANOVA, no specific presentiment model needs to be assumed. However, we point out that this resilience to model type only holds for the ANOVA method as long as the true influence is consistent across stimuli. An inconsistent influence of preceding or succeeding stimuli would be very difficult to isolate with any method, as will be elaborated on somewhat in Discussion.

Table 8 shows how arousal is calculated for each successive trial. The first column in this table contains trial numbers. A trial is here defined as an arousal measurement that is preceded by a stimulus and followed by another stimulus, meaning that the number of trials is one less than the total number of stimuli (the 12 stimuli gives rise to 11 trials). For a given trial, the next column in Table 8 gives the arousal magnitude that has been generated by expectation and cumulated during all previous trials, that is, all trials up to (but not including) the current trial. The third column shows the added arousal magnitudes according to the linear expectation model, and the fourth column the added arousal magnitudes according to the assumed presentiment model, which do not cumulate over trials. In the next column, we have entered an error term, which, for simplicity, has been set to the same absolute value throughout: 0.25, which shifts irregularly between plus and minus (for real data, errors are assumed to be randomly distributed to fit ANOVA). The final arousal values are given in the last column. These values are calculated by adding up all values in the four preceding columns. This is the dependent variable in our two-way ANOVA.

<table>
<thead>
<tr>
<th>Trial number</th>
<th>Previous and subsequent stimulus</th>
<th>Accumulated expectation arousal (b)</th>
<th>Added arousal resulting from expectation (b)</th>
<th>Added arousal resulting from presentiment (c)</th>
<th>Error</th>
<th>Final arousal (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(C) (C)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>(C) (C)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>2.25</td>
</tr>
<tr>
<td>3</td>
<td>(C) (A)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-0.25</td>
<td>3.75</td>
</tr>
<tr>
<td>4</td>
<td>(A) (C)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>(C) (A)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>2.25</td>
</tr>
<tr>
<td>6</td>
<td>(A) (A)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>7</td>
<td>(A) (C)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>8</td>
<td>(C) (C)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>1.25</td>
</tr>
<tr>
<td>9</td>
<td>(C) (C)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>2.25</td>
</tr>
<tr>
<td>10</td>
<td>(C) (A)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td>4.25</td>
</tr>
<tr>
<td>11</td>
<td>(A) (A)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.25</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Note. \(^a\)Arousal generated by expectation of an activating stimulus cumulated over previous trials; \(1\) = one arousal unit, \(0\) = baseline. \(^b\)Arousal increases by one unit in response to presentation of a calming stimulus. \(^c\)Arousal increases by one unit in response to presentiment of an activating stimulus. \(^d\)Obtained by summing the arousal values in the four previous columns.

It should be noted that, for the sake of simplicity, no direct effect of the previous stimulus on the arousal level (for instance, increased heart rate in response to a fear-inducing picture) has been entered into the present experiment. In real experiments, however, such effects do, of course, occur and may in fact be expected to dominate over possible expectation effects. Thus in a real experiment, arousal as a function of the previous stimulus can be
assumed to be a joint effect of direct prior stimulation and expectation. This inability to discriminate between direct arousal effects and additional expectation effects does not reduce the ability of ANOVA to separate expectation effects from presentiment ones, however; it only makes it impossible to detect and identify a possible expectation bias as such.

Table 9 exhibits the three variables used in the present ANOVA, “the final arousal” variable from Table 8 and the two independent variables: the type of preceding and succeeding stimulus, respectively. If there is an expectation bias, a correlation exists between the type of preceding stimulus and arousal level: According to the present expectation model, a calming stimulus, coded as “0,” has the effect of increasing the arousal level by one unit, whereas an activating stimulus, coded as “1,” resets the arousal level to baseline (“0”). If there is a presentiment effect, a correlation exists between the type of succeeding stimulus and the arousal level: According to the present presentiment model, an upcoming activating stimulus has the effect of increasing the arousal level by one unit, whereas an upcoming calming stimulus leaves the arousal level unchanged. Presentiment is also supported by an interaction effect, as an interaction effect occurs when—and only when—a presentiment effect is dependent on the type of preceding stimulus.

Table 10 and Figure 1 show the results of applying an independent two-way ANOVA to the variables shown in Table 9. The significant effect of the type of preceding picture reveals an expectation effect. As the arousal level is higher for calming preceding stimuli than for activating ones, it follows that the expectation effect is consistent with the gambler’s fallacy. (No conclusion regarding the existence or type of expectation effect—whether it acts to increase or decrease the arousal level—can be drawn in real experiments, however, because the direct effect of the previous stimulus on the arousal level could confound any expectation effect; but what you can say—and this is the important thing—is that a possible expectation effect is separated from a possible presentiment effect.) There is also a presentiment effect, as indicated by the significant effect of the type of succeeding stimulus, even though this effect is weaker than the expectation effect. There is no significant interaction effect, but a tendency for the succeeding picture effect to be larger when the preceding picture is calming than when it is activating. Thus, our analysis of these simulated data has revealed a significant presentiment effect that is not dependent on the likewise significant expectation effect (which acts to increase the arousal level in accordance with the gambler’s fallacy.) And although not significant, the interaction effect is consistent with the occurrence of a true presentiment effect, as the interaction effect accounts for the part of the total presentiment effect that is dependent on the type of previous stimulus.

<table>
<thead>
<tr>
<th>Trial number</th>
<th>Dependent variable</th>
<th>Independent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Type of preceding picture</td>
</tr>
<tr>
<td>1</td>
<td>0.75</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2.25</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3.75</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2.25</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.75</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1.25</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2.25</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>4.25</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>1.25</td>
<td>1</td>
</tr>
</tbody>
</table>

*aFinal arousal from Table 8. b1 = activating, 0 = calming*
There are several different ways of applying ANOVA to data from presentiment experiments using more than one participant, which, of course, is common practice. One is to analyse data at the group level from the beginning, thus omitting any analysis at the individual level. This may be done in two different ways. One is to pool data from different participants at the trial-by-trial level, thus treating data as if they were generated by a single person, as illustrated in Table 6. Alternatively, one could apply a between-groups or a within-participant analysis for the whole group of participants, as illustrated in Tables 1 and 2, respectively. Apart from more specific problems associated with these methods (e.g., the occurrence of correlated measures in the pooling method, mentioned previously), the two methods share a more general problem: possible expectation effects may vary considerably in form or strength among participants, perhaps even going in opposite directions in different participants. This raises the possibility that the group results would not accurately reflect the individual results due to the mixture of within- and between-participant effects. This, in turn, means that the group results are not transparent enough to interpret unambiguously. Most notably, we cannot be quite sure that expectation bias has been properly separated from presentiment effects.
An alternative, and in our view better, approach is to apply the following two-step strategy. As a first step, instead of combining the individual data in one way or another, the ANOVA is performed on each individual's trial-by-trial data, as just demonstrated. Thereafter a statistical “counting” procedure is applied: the proportion of individual significant presentiment effects in the group is tested statistically, using, for instance, a binomial test. Thus, if one chooses to perform $N$ individual ANOVAs on the trial-by-trial data (one ANOVA for each participant) and gets $n$ significant main effects of presentiment, the proportion of such effects can be used to determine the statistical significance of presentiment in the given set of participants. For instance, if out of 20 participants, 4 or more show a significant main effect of presentiment, this would suggest that presentiment was occurring at $p < .05$, as indicated by a binomial test. (Also, as suggested earlier, the proportion of significant interaction effects can be used to test for the occurrence of presentiment.)

**Discussion**

All three authors of the present paper have pointed out that presentiment data should be tested for expectation bias. A recent meta-analysis (Mossbridge, Tressoldi, & Utts, 2012) of 26 reports published between 1978 and 2010 indicated that about 70% of the studies satisfied this recommendation by using at least one method of testing for expectation bias—a surprisingly high figure in our minds. More specifically, most people used some variation of Dean Radin’s (2004) approach: looking for a positive trend in the effect as the number of calming stimuli preceding an activating stimulus increases. To our knowledge, none used ANOVA to exclude or factor out expectation bias. We hope tests of expectation bias will soon become still more common than indicated above, and that they even become standard procedure in presentiment experiments, as well as in other, similar experiments (discussed below).

One argument that may be—and has been—raised against worrying a great deal about expectation bias is that this bias, according to simulations (Dalkvist et al., 2002), is generally small—especially when data are pooled before averaging. In these simulations, however, arousal has never been allowed to fall below baseline, which invariably has been set to zero, as in the present paper. However, physiological rebound effects (for example, high heart rate levels being followed by low heart rate levels) would change some zero values to negative values, which would accentuate the expectation bias even further. Accordingly, the occurrence of major biases in simulations using not only positive but also negative arousal values has, in fact, been demonstrated in a recent paper by Kennedy (2013).

However, even if it were true that expectation bias in general does not strongly influence presentiment results, it is critical to examine presentiment data for this bias and to report the results of these examinations. As we mentioned earlier, even a nonsignificant expectation bias can result in a trend that influences the measured presentiment effect. Also, even if expectation biases are generally not strong enough to affect individual studies very much, they may be critical in meta-analyses, where small effects from many studies can add up to give rise to highly significant overall results.

But it is not sufficient that expectation bias is tested for statistically and reported. It is also important that a possible or established bias is handled in such a way that it is totally prevented from affecting measures of a hypothetical presentiment effect.

**A Recommended Two-Step Strategy**

To achieve this goal, a two-step strategy has been recommended in this paper: As a first step, instead of combining the individual data in one way or another, ANOVA is performed on individual trial-by-trial data. As a second step, the proportion of individual significant presentiment effects are tested statistically, using, for instance, a binomial test.

This does not mean, of course, that other methods discussed in this paper can never be useful. Of particular interest as an alternative candidate method is perhaps the single-trial method, as this method can be used to avoid all discussion about expectation bias and still permit mean responses to be calculated to restrict stimulus sampling errors resulting from an unequal number of calming and arousing stimuli. Thus, when the large number of participants that this method requires is available, it could well be used instead of the two-step ANOVA strategy recommended in this paper. Nevertheless, the single-trial method cannot replace ANOVA altogether, for two reasons. One is that few scientists have convenient access to the number of participants required for using this method, even though doing
the experiment at some public place and recruiting a large number of by-passers would be a viable possibility for most researchers. The second reason—the most important one—is the need of a standard statistical method, such as ANOVA, to handle archived data sets.

In a recent paper, Kennedy (2013) proposed a new statistical strategy for handling the expectation bias in presentiment experiments. Instead of using a physiological measure as the dependent variable and the type of forthcoming stimulus as the independent variable, as is customary in presentment research, Kennedy recommended that the two variables exchange roles in a statistical analysis: After appropriate classification, the physiological measure should be used as the independent variable and the type of forthcoming stimulus as the dependent variable. This is in accordance with most traditional precognition experiments (e.g., experiments using RNGs to create future events), where the participant is asked to predict a forthcoming random event. Two major reasons for this recommendation were put forward. One was the fact that no expectation bias will occur when the suggested strategy is used. However, which of the two experimental variables serves as the independent variable and which serves as the dependent variable in a statistical analysis is not of decisive importance for whether or not the bias will appear. It is true that no bias will occur when participants’ predictions serve as the dependent variable. But given that no “forbidden” mean calculations are performed, the bias will also fail to appear when the physiological variable is used as the dependent variable, as was clearly demonstrated earlier in the present paper. Thus, as far as the emergence of the bias is concerned, whether or not any inappropriate mean calculations occur is the crucial factor, not how the two experimental variables are analysed statistically in terms of independent and dependent variables.

The second reason for letting the type of forthcoming stimulus serve as the dependent variable concerns a basic assumption of standard statistical analyses, mentioned before: Measures of the dependent variable are independent of each other. Given that the stimulus order is accurately randomized, this assumption is obviously satisfied when the variable to be predicted by the participant is used as the dependent variable. In contrast, when a physiological measure (or some other response measure) is used as the dependent variable, the assumption of independent measures of the dependent variable can be expected to be violated due to the occurrence of some response pattern. When this happens, the observed alpha level will deviate from the correct one—being too small when correlations among the measures tend to be positive and too large when correlations tend to be negative. There are several ways of dealing with this problem, however. For example, when using ANOVA, as described in the present paper, with a time-lag, the dependent variable may be introduced as a covariate to eliminate, or at least reduce, its auto-correlation, as suggested by Kennedy (2013, p. 246). Alternatively, in the spirit of Kennedy’s (2013) recommended strategy, one could let the physiological variable serve as one of the two independent variables and the type of forthcoming stimulus serve as the dependent variable.

It should be added, however, that the problem of correlated measures of the dependent variable can be expected to diminish considerably in the near future thanks to the “bootstrap” method, a computerized simulation method that, like most other such methods, is free from any statistical assumption (Diaconis & Efron, 1983): This method is currently being implemented by an increasing number of statistical tests in many statistical packages.

Unfortunately, with the exception of the one-trial method discussed above, all of the methods considered in the present paper that avoid the bias altogether—the three methods avoiding the bias in Table 5 (Methods 2-4) and the above-considered method proposed by Kennedy (2013)—share a common flaw: Their power to test for precognition can be expected to be very low, due to the direct effect of the previous stimulus on the physiological variable, which, as noted earlier, can be expected to give rise to a major portion of the error variance. In contrast, by using ANOVA as recommended in the present paper, not only the variance caused by a possible expectation effect, but also the variance reflecting the direct effect of the previous stimulus will be separated from the variance related to the forthcoming stimulus, and the previous stimulus effect will therefore be prevented from disguising any possible precognition effect. Thus, by using ANOVA as described in the present paper, two goals are achieved simultaneously: (a) the expectation bias is eliminated and (b) the error variance is reduced substantially.

One (minor) drawback of the trial-by-trial ANOVA analysis we describe is that it only takes into account bias resulting from the previous trial. However, it would be trivial to add more independent variables into the ANOVA to account for Trials 2 and 3 (previous to the current trial), although this solution still does not consider potential differential effects of particular types of “runs.” For instance, an oscillating sequence of trials (A,C,C,A,C,C,A,C,…) could conceivably produce an expectation for an upcoming calming stimulus (the next item in the repeated series). However, it seems likely to us that such effects will not dominate, as long repeating sequences are generally rare in well-randomized series.
As noted earlier, an inconsistent influence of preceding or succeeding stimuli would be very difficult to isolate, not only with ANOVA but with any method. This raises the question of how serious the possibility of such influences is. Inconsistent influences of preceding stimuli can, in fact, be expected to balance out one another at least to some extent, and may thus be assumed to be relatively harmless. To the extent that inconsistent expectation effects do not balance out, they should be captured by ANOVA as a main effect of the type of previous stimulus or contribute to random errors, rather than lead to bias. Corresponding influences of succeeding stimuli would signal not only presentiment effects going in the expected direction, but also presentiment effects going in the opposite direction (so-called psi-missing). To be sure, such variable presentiment effects would be difficult—or perhaps impossible—to get at, but there is no reason to think they occur very often.

A comment should also be made about our recommendation to perform individual ANOVAs. It may be argued that a drawback of this procedure, as compared with combining individual data before applying ANOVA, is the expected reduction of power. How significant the expected reduction will be is hard to tell a priori and should be assessed in future research, ideally by means of simulation. We think, however, that a reasonable reduction in power is a price worth paying to maintain the transparency of the analysis. Nevertheless, if the individual data sets are very small, combining individual data in one way or another before ANOVA is applied may be necessary.

Generalizations

It should be noted that the gambler’s fallacy is not the only cognitive bias that can create a response bias that could influence presentiment measures. An opposing bias, the “hot-hand” bias, has been described elsewhere (e.g., Roney & Trick, 2009). The hot-hand bias applies primarily to people who believe that a sentient agent controls the outcomes of the events in a series, for instance points in a basketball game. A participant who is subject to the hot-hand bias would expect a series of activating images to continue. However, it is not clear whether this bias would produce an increase in the type of expectation bias effect examined here (if arousal continues to increase before each activating image), or if it would produce an effect in the opposite direction (if arousal drops as each expected activating image is presented). Whether participants hold the hot-hand bias, the gambler’s fallacy bias, or neither, any bias that is present can be separated from an existing presentiment effect using the trial-by-trial ANOVA analysis described here.

Contrary to what we have argued elsewhere (Dalkvist, Mossbridge, & Westerlund, 2013), and based on remarks in Kennedy (2013), we believe that the present bias has not affected previous conscious precognition experiments very much, if at all, and that it will not be a threat to such future experiments. The reason: Just as in presentiment experiments, in standard conscious precognition studies, where participants are required to predict forthcoming events, no bias will occur unless means are calculated that assign different weights to responses depending on the type of forthcoming event. Such mean calculations are unlikely to occur in a standard precognition experiment. They could occur, though, at least in theory. It is thus possible to create a bias in a conscious precognition experiment (e.g., an experiment using RNG) by treating data in the same way as presentiment data are treated to give rise to a bias. But as this type of analysis is far-fetched (in contrast to just correlating predictions with outcomes, for example), we do think it is rare, or has perhaps never occurred. Nevertheless, inspection of the literature to discover whether this belief is true should be undertaken.

But what about mainstream science? Outside parapsychology, there are a number of psychophysiology studies that formally resemble presentiment experiments as conceptualized in the present paper, and such studies are therefore potentially vulnerable to expectation bias. A few of them have been analysed with the aim of finding evidence of anticipatory effects (Bierman, 2000; Mossbridge et al., 2012). Together, these analyses do point to the occurrence of such effects, but it is still an open question whether these possible anticipatory effects resulted from expectation bias as conceptualized in the present paper. There are also other research areas, particularly within psychology (for instance, the areas of memory and attention), where expectation bias might occur. However, in mainstream psychology, randomization without replacement, which, in combination with feedback, allows the participant to predict future stimuli based on assumptions about how stimuli are distributed, seems to us to be more common than randomization with replacement, which does not permit any such predictions. This means that, on top of the expectation effect considered here, which may properly be regarded as numerical bias, attributable to inappropriate mean calculations, rather than to expectations per se, there is another, ”real,” expectation effect. The combination of the two expectation effects may, conceivably, be quite dramatic.
However, from the perspective of mainstream science, the critical question is not how to prevent a possible expectation effect from affecting measurement of a possible presentiment effect, but instead how to prevent expectation effects from affecting measurement of the direct effect of the previous stimulus (for instance, a possible physiological effect of frightening pictures). This is a tricky problem, as the direct effect of the previous stimulus is confounded by the expectation effect due to their common triggering stimulus. A possible solution to this problem may be to take advantage of some possible difference in time between the direct effect of the previous stimulus and the corresponding expectation effect (conceivably, the expectation effect may be somewhat delayed compared to the direct effect). In any case, the particular ANOVA recommended in the present paper is not applicable to mainstream data, except, of course, for attempts to demonstrate or control for presentiment effects in such data.

Concluding Remarks

During the short history of experimental parapsychology, it has commonly been assumed that any properly randomized stimulus/response experiment will be free from systematic errors. This assumption—based on the plausible, but wrong, belief that in such experiments, finite stimulus sequences behave just as infinite ones do—has been especially important in presentiment research, where responses to previous stimuli ideally should be totally prevented from affecting hypothetical responses to future stimuli. The discovery of the present expectation bias has revealed the myth inherent in the assumption that a properly randomized stimulus/response experiment is free from error. An obvious practical implication of this revelation is that parapsychologists have to guard very carefully against this bias so as not to confuse presentiment with artifacts. As a bonus, possible presentiment effects may be more easily unveiled by controlling for the expectation effect. Hopefully, the present paper will be of some help in this endeavour. In particular, we hope that the two-step strategy recommended here will serve as a useful safeguard against mistaking biased predictions for presentiment.

References


Department of Psychology
Stockholm University, S-106 91
Stockholm, Sweden
jandalkvist@gmail.com
Acknowledgement

We want to thank the reviewers for useful comments that resulted in extensive revisions and improvements from earlier versions.

Abstracts in Other Languages

German

Wie läßt sich der Einfluss der Verzerrung der Erwartungsbildung bei Presentiment- und ähnlichen Experimenten beseitigen?


Spanish

Cómo quitar la influencia del sesgo de expectativa en experimentos de presentimiento y similares: Una estrategia recomendada

Resumen: Reconsideraremos el sesgo de expectativa en los llamados experimentos de presentimiento, con especial atención en la forma de cómo manejar la situación. En tales experimentos se piensa por lo general que el presentimiento se ha demostrado al mostrar que diferencias fisiológicas significativas preceden a estímulos que presumiblemente dan lugar a diferentes niveles de excitación. A menudo, estas diferencias sugieren que la activación fisiológica más probablemente precede a los estímulos excitantes que a los tranquilos. Es conceivable, sin embargo, que tales reacciones puedan explicarse como el resultado del sesgo de expectativa de la falacia del jugador. Este sesgo se basa en la (falsa) noción de que la probabilidad de un estímulo excitante crece a medida que el número de estímulos tranquilos aumenta. Discutimos diferentes formas de controlar o evitar el sesgo. Nuestra recomendación resultante es utilizar el análisis de varianza (ANOVA) para separar el efecto del sesgo del efecto hipotético de presentimiento, preferentemente a nivel de prueba por prueba. También se recomienda la aplicación de ANOVA a cada participante por separado y usar un método de “conteo” para evaluar los posibles efectos de presentimiento a nivel
Deacribimos la aplicación de ANOVA con un ejemplo simulado. Anticipamos que ANOVA puede controlar no sólo el sesgo falaz del jugador falacia sino también sesgos similares, tanto en experimentos de presentimiento como en experimentos de precognición conscientes.

French

COMMENT EVITER L’INFLUENCE DU BIAIS D’ATTENTE DANS LA RECHERCHE SUR LE PRESSEN- TIMENT ET DANS DES EXPERIENCES SIMILAIRES : RECOMMENDATIONS STRATEGIQUES

RESUME : Nous examinons ici le biais d’attente dans les expérimentations sur ce qu’on a appelé le pressentiment, et surtout sur la manière de l’éviter. Dans ce type de test, on tente de démontrer la présence de pressentiment en montrant que des différences physiologiques significatives précèdent les stimuli censés produire différents niveaux de réaction. Ces différences suggèrent souvent que la stimulation physiologique semble anticiper sur l’apparition des stimuli lorsque les stimuli sont excitants plutôt que calmes. Toutefois, il est évident que de telles réactions pourraient aussi être expliquées comme le résultat d’un biais d’attente du type du sophisme du joueur. Ce biais se base sur la (fausse) notion selon laquelle la probabilité que le prochain stimulus soit un stimulus excitant augmente en fonction du nombre de stimuli calmes consécutifs. Différentes manières de contrôler ou d’éviter ce biais sont discutées. La recommandation qui en résulte consiste à utiliser l’analyse de la variance (ANOVA) pour séparer l’effet du biais d’un effet hypothétique de pressentiment, de préférence à un niveau d’essai par essai. Nous recommandons également d’appliquer l’ANOVA pour chaque participant séparément et d’utiliser une méthode de « comptage » pour tester de possibles effets de pressentiment au niveau du groupe. L’application de l’ANOVA est illustré en employant un exemple simulé. Nous pensons que l’ANOVA pourra non seulement régler le biais du sophisme du joueur mais également d’autres biais similaires, à la fois dans les expérimentations sur le pressentiment et dans certaines expérimentations sur la précognition consciente.
ABSTRACT: Since the 19th century many psychiatrists and psychologists have considered mediumship to be related to the subconscious mind and to dissociative processes produced mainly by internal conventional processes of the medium’s mind. However, some psychologists and psychical researchers active between the last decades of the 19th century and the 1920s expressed a different view. Individuals such as Théodore Flournoy, Cesare Lombroso, Enrico Morselli, Frederic W. H. Myers, Julian Ochorowicz, Charles Richet, Eleanor Sidgwick, and Eduard von Hartmann, argued that some mediums combined dissociation with supernormal phenomena such as knowledge acquired without the use of the senses, and the production of physical effects seemingly beyond the normal bodily capabilities. Depending on the theorist, other issues such as pathology and discarnate agency were also part of the discussions. The supernormal was never accepted by science at large and today is rarely mentioned in the dissociation literature. But ideas related to the supernormal were part of this literature. A complete history of dissociation, and of the subconscious mind, should include consideration of this body of work.

Keywords: subconscious mind, dissociation, history of mediumship, psychical research, telepathy

In his book *The Discovery of the Unconscious* Ellenberger (1970, pp. 317–318) argued that by 1900 four main functions of the unconscious mind had been established. These functions were conservative (the repository of memories and perceptions), dissolutive (automatic and dissociative aspects that may interfere with normal functions), creative, and mythopoetic (or fictional fabrications of all sorts). In this paper I will focus on one function not included by Ellenberger, what we would refer to today as “parapsychological.” These phenomena were also referred to as psychic, metapsychic, and supernormal. My discussion will focus on theoretical ideas about mediumship and the subconscious mind that include the supernormal, ideas prevalent from the late 19th century to late in the 20th century. The purpose of the paper is to reacquaint contemporary students of dissociation and of mediumship with this nearly forgotten past, a past that at the time interacted with and affected contemporary psychological and psychiatric work toward constructing the concept of dissociation and the idea of hidden powers of the subconscious mind.

To this day the term dissociation has different meanings and conceptualizations (Nijenhuis & Van der Hart, 2011; Spitzer, Barnow, Freyberger, & Joergen, 2006), but it is frequently used to refer to a misunderstood process underlying disruption or separation of memory, identity and sensations from consciousness. The term was not used by the authors discussed in this paper but they evidently believed in such hypothetical process to refer not only to the equally ill-defined mediumistic trance, but to the various manifestations of hypnosis, and hysteria, the latter which some postulated included fugues, amnesia, and so-called “secondary” (or more than two) personalities.

Although I (Alvarado, 2002, 2010a) and several others (e.g., Crabtree, 1993; Ellenberger, 1970; Plas, 2000; Shamdasani, 1994) have argued for contributions to the study of psychic phenomena or the supernormal, to psychology and psychiatry, and particularly to ideas about dissociation and concepts about nonconscious levels of the mind, here I focus on ideas that have not received much attention.

Mediumship in Context: Spiritualism, Psychical Research, and Psychiatry

The psychical researchers of the 19th century inherited from previous movements the idea that human beings had powers that could transcend physical limitations (for overviews see Alvarado, 2012; Crabtree, 1993; Inglis, 1992; Méheust, 1999a; Podmore, 1902a, 1902b). The work of the mesmerists—as seen in books such as *Natural and Mesmeric Clairvoyance* (Esdaile, 1852)— promoted the view that some people could influence others at a distance to produce phenomena such as cures and thought transference. In addition to publicizing psychic
phenomena, thus creating an intellectual and experiential context for the unorthodox, mesmerism was important for the development of Spiritualism and mediumship in various ways. Among them were the various discussions in the mesmeric literature on nonphysical ideas (Alvarado, 2012, pp. 39–40) and the “mediumistic-like” phenomena reported by some mesmerized individuals (Crabtree, 1993, pp. 196–212).

As social reformer Robert Dale Owen (1860), who lived from 1801–1877, discussed in Footfalls on the Boundary of Another World, Spiritualists assumed that some individuals had the ability to perceive and to channel phenomena caused by spirits of the dead. Psychical research emerged as an attempt to continue and systematize the work of the mesmerists and the Spiritualists by exploring hidden dimensions of human functioning and the possibility of spirit agency (Gauld, 1968; Lachapelle, 2011; Moore, 1977; Wolffram, 2009).

While most contemporary scientists seemed to accept a closed model of the mind in which the workings of the subconscious and the phenomena of dissociation were explained by intrapsychic psychological, physiological, and medical factors (including outside influences such as suggestion and trauma), some of the psychical researchers (like the Spiritualists) argued for a model in which the mind was an open system not bound by the limitations of the nervous system and the senses (Gauld, 1968). Although the work in London of the Society for Psychical Research (SPR), founded in 1882, focused on different supernormal phenomena, such as spontaneous and experimental telepathy (e.g., Gurney, Myers, & Podmore, 1886; Guthrie, 1884; see also Gauld, 1968; Le Maléfan, 2008; Luckhurst, 2002), one of the major lines of work that the SPR and researchers outside England conducted was the study of mediumship.

Several historians have chronicled the prominence of mediumship in the 19th century. Some examples are historical studies about spiritualism and mediumship in England (Owen, 1990), Italy (Biondi, 1988), and the United States (Braude, 2001) (see also Alvarado 2010a; Alvarado and Zingrone, 2012, Edelman, 1985, and Galvan, 2010). Some mediums, such as Emma Hardinge Britten (1823–1899) and Cora L.V. Tappan (later Richmond; 1840–1923), were trance or inspirational speakers on social, scientific and philosophical topics (Hardinge, n.d.; Tappan, 1875). Many communications recorded through mediums suggested to some, as one historian wrote, that mediums were in “an elevated state, providing access to spirits and therefore to knowledge of the world beyond inaccessible to conscious human beings” (Braude, 2001, p. 87). The early American spiritualist literature contained many spirit communications about the nature of the afterlife and moral and philosophical topics (e.g., Hammond, 1866; Hare, 1855). Many recorded communications from famous individuals, among them Francis Bacon (1561–1626) and Emanuel Swedenborg (1688–1772) (Edmonds & Dexter, 1853), Thomas Paine (1737–1809; Hammond, 1866), and American presidents George Washington (1732–1799) and Thomas Jefferson (1743–1826; Spicer, 1853). Some, like Allan Kardec (pseudonym of Hippolyte Léon Rival Denizard, 1804–1869) in France, questioned the spirits through mental mediums (Kardec, 1857). The “spirits” dictated teachings about such issues as God, the spiritual world, reincarnation, the intervention of spirits in life, moral laws dealing with various issues (e.g., society, progress, freedom), evocation of spirits, and mediumship, among many other topics (see also Edelman, 1995; Sharp, 2006). But there were also more specific communications in which mediums presented verifiable information about deceased individuals (e.g., Moses, 1879).

Among the influential phenomena of mediumship were verbal communications, automatic writing (including the production of literary works), as well as visions (e.g., Edmonds & Dexter, 1853; Moses, 1879). Many were impressed by what one author called personification, or “the appropriation or adoption of the name and many times of the character of a personality foreign to that of the medium” (Aksakof, 1890/n.d., p. 30; this and other translations are mine). Such phenomena, reported through the history of Spiritualism and psychical research with many mediums, received much discussion (see Alvarado, 2011b).

In addition to the mental mediums, there were physical mediums thought to manifest spirits through physical manifestations. These phenomena included acoustic ones such as raps (sounds produced on objects) or voices, the movement of objects, luminous effects, the appearance of writing on slates or paper, and materializations of figures, both solid and misty, that supposedly represented the spirits of the dead. Two famous examples were mediums Florence Cook (1856–1904), well-known for her materializations of a “spirit” called Katie King (Corner, Corner, Luxmoore, Tapp, & Harrison, 1873), and D. D. Home (1833–1886), in whose presence sitters reported levitations of tables and of the medium, remarkable beautiful materializations of hands, and immunity to fire, among many other manifestations (Adare, 1869).

From the beginning of mediumship there were many attempts to explain the phenomena in conventional terms, that is, in terms of the closed model of the mind described above. Other than fraud, there were a variety of medical and psychophysiological explanations of mediumship (Alvarado & Zingrone, 2012; Crabtree, 1993; Le Maléfan, 1999; Owen, 1990). These included the view that some mediumistic phenomena represented examples of...
of automatic activities of the nervous system (e.g., Faraday, 1853). Others argued that trances frequently were a disease of the nervous system “in which the cerebral activity is concentrated in some limited region of the brain, the activity of the rest of the brain being for the time suspended” (Beard, 1879, p. 67).

Many physicians wrote about mediumship, particularly mental mediumship, as manifestations of psychopathology. (On mediumship and pathology, see Alvarado and Zingrone, 2012; Le Maléfan, 1999; Moreira-Almeida, Almeida, and Lotufo Neto, 2005; and Owen, 1990.) These works included ideas of uterine pathology (Marvin, 1874). Pierre Janet (1859–1947) wrote that mediumship was related to “a particular morbid state analogous to those that may later become hysteria or insanity: mediumship is a symptom and not a cause” (Janet, 1889, p. 406).

The medical interpretation of mediumship was reinforced by evidence of personality changes that resembled spirit communications in contexts other than mediumship. These included spontaneous changes of personality (e.g., Elliotson, 1846), among them the famous Félida X case (Azam, 1887), as well as seeming secondary personalities that appeared during the hypnotic trance (e.g., Janet, 1889; Lang, 1843), or as the result of suggestion. This is what one researcher referred to as “objectification of types” (Richet, 1883).

The production of these dramas under hypnotic suggestion opened the door to later psychological speculations about mediumship, such as those assuming that some spirit communications were the “product of the medium’s subconscious imagination, working from recollections and latent worries” (Flournoy, 1899, p. 144). This and later analyses of the mediumistic romances of a previous life in India and France and of life on Mars produced by the medium Hélène Smith (pseudonym of Catherine Elise Muller, 1861–1929; Flournoy, 1900) supported one of the main assumptions of the medical and psychological model: Mediumistic phenomena were produced by the internal mental resources of the medium, perhaps influenced by suggestion. This is what one writer referred to as the “enormous suggestibility and autosuggestibility of mediums” (Flournoy, 1900, p. 443).

However, there were others who argued that the subconscious mind had more than psychological and pathological functions in mediumship. These individuals proposed the concept of supernormal powers of the subconscious mind to explain mediumistic phenomena. Although ideas along these lines were present earlier during the 19th century (Alvarado, Nahm, & Sommers, 2012; Crabtree, 1993), my focus will be examples published between the late 19th century and the 1920s. I will start with the work of two important and influential pioneers, and will follow with the work of others to account for specific cases of mediumship. My summary of relevant ideas will be confined to a small number of students of the subject, and the reader should be aware that my discussion is not exhaustive.

The Influential Writings of Frederic W. H. Myers and Eduard Von Hartmann

An important early theoretician who influenced thinkers such as Pierre Janet and William James (1842–1910) was Frederic W. H. Myers (1843–1901). Myers—a classical scholar, psychologist and psychical researcher—conducted theoretical work about the subliminal (subconscious) mind and the supernormal in the context of the early SPR. In his view, “the stream of consciousness in which we habitually live is not the only consciousness which exists in connection with our organism. Our habitual or empirical consciousness may consist of a mere selection from a multitude of thoughts and sensations, of which some at least are equally conscious with those that we empirically know” (Myers, 1892a, p. 301). Myers assumed that a subliminal self manifested in disintegrations of personality as well as in more positive phenomena such as creativity and mediumship. In addition, he believed that the mind was in charge of the body and independent of it, a view different from many of his contemporaries who followed physiological and pathological assumptions (Crabtree, 2003).

Myers argued that mediumship and other phenomena such as creativity and telepathy were expressions of the subliminal mind. This took place largely through sensory and motor automatisms (Myers, 1892b, 1893). Sensory automatisms included visions, voices, dreams, and intuitions, whereas motor automatisms included phenomena such as automatic writing and table turning as well as trance speaking. Both types of automatisms were, according to Myers, conveyors of messages. In his view, “the message which they bring comes from a stratum or phase of our own personality which is chaotic and fragmentary . . . . a message conveyed by automatic action from the subconscious to the conscious or waking self . . . .” (Myers, 1889, p. 535).

In an early paper Myers (1884) argued that many mediumistic communications could be explained without assuming any supernormal ability. For example, automatically produced scripts seemingly done without the conscious knowledge of the automatist could take place, Myers suggested, at will or through low- or high-level unconscious guidance from the automatist’s mind. However, Myers also mentioned two other possible causes of these communications that do involve supernormal function: telepathy from the sitter and the influence of discarnate
beings. Although he did not discuss the spirit hypothesis in any detail, Myers stated that a few cases suggested thought transference from the sitters attending the séance. This idea was further discussed in a later paper, where he suggested that during motor automatisms a secondary self had a “purposive activity of its own” (Myers, 1885, p. 27). Myers further wrote: “It is the secondary self . . . which receives or recognises the telepathic impact . . . and in some way or other furnishes an intelligent reply” (Myers, 1885, p. 28).

Later Myers discussed the action of spirits of the dead. He argued for “contact between the discarnate mind and the brain of the automatist . . . .” (Myers, 1903, p. 196). The spirit communicator, he believed, used the “capacities” of the medium’s brain to communicate, an idea not new with Myers (cf. Kardec, 1863, p. 281). However, in Myers’ view the communication process “must be limited by the idiosyncracies of the medium” (Myers, 1903, p. 249). Communications, he argued, generally presented a mix between influences from the medium’s subliminal mind and a discarnate agency.

Another influential theoretician in Europe was German philosopher Eduard von Hartmann (1842–1906; see Treitel, 2004, pp. 13–15, 21–22; Wolffram 2009, pp. 41–43). In his view, a medium was a person capable of entering a somnambulistic state either spontaneously or through some induction procedure. Mediums showed “a certain disorganization of the nervous system; that is, the lower and the middle nerve centres are too independent of the highest, reflex-prohibiting centre of consciousness self control; they are, in other words, . . . hysterical . . . .” (Von Hartmann, 1885, p. 406).

According to von Hartmann, mediums showed two types of somnambulism: manifest and masked. In manifest somnambulism mediums had no recollection of the phenomena they produced. The masked type implied that the waking consciousness during mediumistic phenomena had no awareness of the action of a secondary consciousness. Like others before him (for a review see Crabtree, 1993), von Hartmann believed that some of the automatic manifestations of mediumship, such as automatic writing, could take place without any guiding intelligence, being the action of nervous mechanisms without consciousness. But many other manifestations showed a “somnambulistic consciousness accompanying these mechanical material brain processes and enlightening them with real intelligence” (p. 420).

Von Hartmann explained the mediumistic tendency towards symbolization and personifications as the simultaneous action of waking consciousness and enhanced memory access, as well as thought transference and clairvoyance. All of this worked together with the “symbolizing and personifying tendency of the somnambulistic consciousness . . . to put these communications in the mouth of an absent person, its dramatic metamorphosing talent must at the same time, succeed in dressing out the communication with all sorts of small external traits appropriate to the person represented” (p. 453).

Von Hartmann also argued that the somnambulistic consciousness could manage nervous forces that, when exteriorized from the body, could cause mediumistic physical phenomena. As he wrote:

This liberation of directive radiation of nerve force is under all circumstances . . . not a function of those parts of the brain which serve as support to the conscious will, but of deeper-lying layers of the brain which
either coincide with those supporting the somnambulistic consciousness, or are more approximate to them than to the first. It is therefore, no wonder that the development of a magnetic-mediumistic nerve force is stronger in the somnambulistic than in the waking state . . . . (p. 442).

These ideas are part of an old tradition of ideas of “magnetic,” “nervous,” or “psychic” forces related to the workings of the human body and considered to be the agent behind the physical phenomena of mediumship (Alvarado, 2006, 2008; Alvarado, Nahm, & Sommer, 2012; Crabtree, 1993).

Speculations on the Mediumship of Leonora E. Piper

In the mid 1880s psychologist and philosopher William James (1842–1910) had an experience with the American medium Leonora E. Piper (1857–1950) that led him to write that “taking everything that I know of Mrs. P. into account, the result is to make me feel as absolutely certain as I am of any personal fact in the world that she knows things in her trances which she cannot possibly have heard in her waking state, and that the definitive philosophy of her trances is yet to be found” (James, 1890b, pp. 658–659). Mrs. Piper was the first mental medium studied under controlled conditions for a prolonged period of time. Many researchers conducted studies with her during the late 19th century.

Mrs. Piper’s trance phenomena passed through different stages over time, changing or adding spirit controls including the well-known “Dr. Phinuit” and the George Pelham (really called George Pellew) controls. She communicated through automatic writing and trance speaking. In trance speaking, Piper’s spirit control relayed information from a supposed spirit, sometimes answering questions posed by séance participants. In addition, sometimes she seemed to be possessed by the communicating spirit. These performances convinced others besides James that she acquired information beyond the reach of the senses.

Following James, psychical researcher Richard Hodgson (1855–1905) conducted the first systematic investigations of Piper’s trance phenomena (see Berger, 1988, pp. 11–33). The following was written by a sitter, Reverend and psychic investigator Minot J. Savage (1841–1918), as reported in Hodgson’s (1892) first report showing personation and apparent veridical information. The doctor mentioned was “Dr. Phinuit,” the medium’s spirit control at that time:

Immediately on becoming entranced, her control, Dr. Phinuit, said there were many spirit friends present. Among them he said was an old man, whom he described, but only in a general way. Then he said, “He is your father, and he calls you Judson.” Attention was called to the fact that he had a peculiar bare spot on his head, and Mrs Piper put her hand on the corresponding place on her own head . . . . My father had died during the preceding summer . . . He wasn’t at all bald, but when quite young had been burned; so that there was a bare spot on the right side of the top of his head, perhaps an inch wide and three inches long, running from the forehead back towards the crown . . . . This was the spot that Mrs. Piper indicated. Now as to the name by which he addressed me: I was given the middle name, Judson . . . . father always used, when I was a boy, to call me Judson, though all the rest of the family called me by my first name, Minot. (Hodgson, 1892, p. 100)

In his first report Hodgson (1892) stated his belief that Piper’s trance was of the “type where the change to the trance personality involves a partial obliteration of the facts known to the normal waking self” (pp. 55–56). He did not believe that there was enough evidence to support the discarnate agency hypothesis, although he said at the end of the paper that further developments in the Piper mediumship were suggestive of this. Nonetheless, Hodgson concluded that the trances showed “a large residuum to be attributed to some supernormal faculty” (p. 9). Regarding his first six sittings with Piper, Hodgson thought that the medium “had access to portions of my ‘subconscious’ mind” (p. 10). Still, Hodgson did not think telepathy could explain all the veridical communications. As mentioned below, he later stated his belief in spirit communications.

Another early Piper researcher, classical scholar and banker Walter Leaf (1852–1927; see Anonymous, 1927), argued that in the medium’s “abnormal state there is a quite exceptional power of reading the contents of the minds of sitters; but that this power is far from complete” (Leaf, 1890, p. 567). The thought transference process suggested by Leaf was one related to sitters’ subconscious minds, that is, content not consciously recollected at the time of the séance. This gave the impression that a spirit was communicating. Others, such as American philosopher
William R. Newbold (1865–1926; see Sidgwick, 1926), doubted that Piper’s telepathy and clairvoyance could be organized by the medium’s mind into a “mosaic of thought, which . . . often irresistibly suggests the habits, tastes, and memories of some friend deceased . . . .” (Newbold, 1898, p. 9).

Frank Podmore (1855–1910; see Dingwall, 1963), an SPR member well-known for his skepticism regarding physical phenomena, was impressed by Piper. In an early statement Podmore (1897) said that the medium presented evidence of “something beyond telepathy” (p. 454). He did not clearly state his belief in the supernormal in a later article (Podmore, 1898), but in his book *Modern Spiritualism* expressed his conviction that Piper’s trance utterances provided evidence of supernormal perception (Podmore, 1902b, p. 342). Trance personalities, he concluded, were “illustrations of the plastic powers of the medium’s own spirit, rather than as representing alien intelligences” (p. 343). Podmore (1908) argued that it was clear that Piper’s changes of personality were shaped in part by suggestion but those suggestions could be psychically conveyed, “telepathically filched” (p. 329) from the sitters (see also Podmore, 1910).

Educator and once principal of Newnham College Eleanor Sidgwick (1845–1936; see E. Sidgwick, 1938) authored an impressive psychological study of Mrs. Piper that also addressed telepathy. As she concluded regarding Mrs. Piper’s trance:

> I think it is probably a state of self-induced hypnosis in which her hypnotic self personates different characters either consciously and deliberately, or unconsciously and believing herself to be the person she represents, and sometimes probably in a state of consciousness intermediate between the two. In the trance state her normal powers transcend in some directions those of her ordinary waking self . . . . And further — what makes her case of great importance — she can obtain, imperfectly and for the most part fragmentarily, telepathic impressions. (Sidgwick, 1915, p. 330)

More than this, Sidgwick argued that telepathy could provide the “material necessary to successful personation” (p. 319). This assumed that a dissociative process (the trance and the personation accompanying it) could incorporate telepathic information.

In Germany, philosopher Traugott Konstantin Oesterreich (1880–1949; see Melton, 2001, p. 1139) was skeptical of the spirit explanation of mediumship. He believed that the subconscious creative imagination of Mrs. Piper was active during her trance, a process “of daily occurrence in modern occultism by reason of traditions and beliefs which are passed on from one medium to another” (Oesterreich, 1923, p. 49). But in his view Piper’s subconscious imaginative potential consisted of more.

Oesterreich suggested that Mrs. Piper’s veridical communications involved “an elaboration by the creative imagination of Mrs. Piper’s telepathically-acquired knowledge and by her telepathic faculty working in conjunction with the minds of others . . . .” (pp. 44–45). Echoing Leaf, he argued further that the medium’s telepathy could tap into the latent thoughts of living people.

Others defended supernormal explanations that related to the spirits of the dead. In his second report on Piper, Hodgson (1898) became convinced that the supernormal content of Mrs. Piper’s trances required spirit intervention. Some spirit communicators were more clear than others; Hodgson believed that this should not be the case if thought transference was in operation. In addition, confused communications suggested to Hodgson that they reflected the communicator’s illness or other relevant conditions at the time of death. In Italy and in the United States, respectively, others defended the idea of spirit agency to explain Mrs. Piper’s veridical trance phenomena (Bozzano, 1926; Hyslop, 1901).

William James, on the other hand, did not seem so sure of the final explanation of the phenomena produced by Mrs. Piper. Studying the Piper communications that allegedly came from the deceased Richard Hodgson, James (1909) was willing to consider their content supernormal. He wrote: “The active cause of the communications is on any hypothesis a will of some kind, be it the will of R. H.’s spirit, of lower supernatural intelligences, or of Mrs. Piper’s subliminal . . . . yet the major part of it is suggestive of something quite different—as if a will were there, but a will to say something which the machinery fails to bring through” (p. 116). This “will,” he suggested, could “tap,” possibly the sitter’s memories, possibly those of distant human beings, possibly some cosmic reservoir in which the memories of earth are stored, whether in the shape of “spirits” or not. If this were the only will concerned in the performance, the phenomenon would be humbug pure and simple, and the minds
tapped telepathically in it would play an entirely passive rôle—that is, the telepathic data would be fished out by the personating will, not forced upon it by desires to communicate, acting externally to itself. (pp. 116–117)

James expressed the belief that there were traces of a will to communicate coming from Hodgson and that Mrs. Piper’s telepathy was not enough to explain the results. But he was willing to speculate on the interactions of many factors. For example, James argued that there could be combinations of two different wills. One was a “will to personate,” and another the will to communicate, the latter coming from a discarnate source (p. 117). Furthermore, he also speculated on the interaction of the sitter, the medium and the communicating spirit: “The sitter, with his desire to receive, forms, so to speak, a drainage-opening or sink; the medium, with her desire to personate, yields the nearest lying material to be drained off; while the spirit desiring to communicate is shown the way by the current set up, and swells the latter by its own contributions” (p. 120).

Speculations on the Mediumship of Eusapia Palladino

Eusapia Palladino (1854–1918), a well known Italian physical medium, has been generally considered to have played an important role in the development of psychical research. Not only did her performances contribute to the development of ideas about mediumship, but she also was instrumental in convincing many researchers that phenomena such as telekinesis and materializations were real (Alvarado, 1993, 2011a; Carrington, 1909).

During a séance Palladino usually sat on one side of a table around which several sitters were posted. Behind her was a “cabinet” formed of a corner or area behind a curtain where objects were sometimes placed to be affected by telekinesis. The following is a typical example:

The séance began in full light, and whilst the medium was still quite conscious, movements at first slight, then stronger, began in the séance table, which raised three of its feet. In full light the slight outward movements of the curtain on the left were observed. The medium asked by means of the table (five raps) that the light might be lessened; this was done rather slowly, and the strong red light . . . fell directly on to the medium’s eyes, which occasioned in her a fit of hysteria; she wept and cried out as if demented, hitting her face repeatedly with her fists. This was a genuine fit of hysteria . . . .

When the attack was over, Eusapia was no longer in her normal state of consciousness, and no longer spoke in the first person; she spoke as if she were John King . . . . In the meanwhile, the table on which the toys had been placed, and which we call No. 1, made a noise in the interior of the cabinet, from which it at last came out completely. Then there began to arrive on the séance table many objects from table No. 1: a sheet of paper, a little wooden sheep and a mandoline; the latter was accompanied by the curtain which covered the handle; the curtain . . . came back and covered the handle of the mandoline, and a hand, which was not that of the medium or of the sitters, pulled the hair of the person who had pushed back the curtain. At the same time we heard a scratching on the strings of the mandoline. (Aggazzotti, Foà, Foà, & Herlitzka, 1907, pp. 366–368)

Phenomena of this sort were frequently observed by Palladino’s investigators, many of whom were convinced of the genuineness of the phenomena while recognizing the medium committed fraud frequently. Three of these researchers were Italian psychiatrist and criminologist Cesare Lombroso (1836–1909), Polish psychologist Julian Ochorowicz (1850–1917), and Italian psychiatrist Enrico Morselli (1852–1929; see Domanski, 2003, Guarneri, 1988, Knepper & Ystehede, 2013).

In Lombroso’s (1892) view, Palladino could produce telekinesis and other phenomena while entranced but she was also a hysterical that showed signs of epilepsy. In his view hysterics and hypnotized individuals had some nervous centers activated at the expense of others that were paralyzed or arrested. This allowed for a transposition and transmission of psychic forces, which in entranced mediums could be exteriorized from the body as a motor force to cause telekinesis and other physical phenomena.

In a later book, translated into English as After Death—What?, Lombroso (1909) classified Palladino’s phenomena in relation to the medium’s mental state. In his experience “the rarer and more important of the phenomena (for instance, apparitions of phantasms), so much the heavier is the trance of the medium” (p. 120). Eventually,
Lombroso (1908, 1909) argued that some phenomena were produced by spirits who used the medium’s vital force to accomplish such feats as the medium’s levitation. In addition to Lombroso, other researchers argued that some of Palladino’s phenomena, particularly full body materializations, could not be explained solely by the medium’s vital force directed by her subconscious mind (e.g., Carrington, 1909; Venzano, 1907).

Ochorowicz (1896; see also Alvarado, 2010b) did not accept spirit intervention to account for Palladino’s phenomena but did accept the force model. He divided the manifestations of mediums into inferior and superior ones using Palladino as an illustrative case. Inferior manifestations were those consisting solely of fraud and automatism; superior manifestations were those produced beyond the confines of the body. Most mediums used fraud frequently due to lax controls and the expectations of the circle. Fraud, according to Ochorowicz, “dominates their repertoire and the habits of their nervous systems” (p. 111).

In Ochorowicz’s view, at the beginning of a séance it was important to give the medium time to go through phases of physiological “doubling” or dissociation that would eventually culminate in phenomena beyond conventional dissociation. This doubling referred to unspecified brain centers of automatic action, as well as to consciousness in relation to somnambulistic ideas and autosuggestion. But Ochorowicz was clear that it was possible for Palladino to exteriorize a force from her body to produce telekinesis and materializations and that the process could involve the forces of other sitters as well. This view was discussed in a later publication unrelated to Palladino. Ochorowicz discussed a new and revolutionary form of dissociation shown by mediums. This was a process that took place “between organs and the dynamic principle that animates them; which can go as far as extra-organic manifestations, that is . . . out of the body” (Ochorowicz, 1909, p. 757).

Enrico Morselli, whose book *Psicologia e “Spiritismo”* (Morselli, 1908a, 1908b) was almost completely dedicated to Palladino’s mediumship, also contributed important ideas. Like so many before him, Morselli argued that mediumship was “an abnormal fact of human physio-psychic personality . . .” (Morselli, 1908a, p. 95). But he became fully convinced that Palladino produced physical effects such as telekinesis and materializations. According to Morselli (1908a) the low intellectual content of these effects was indicative of psychological disaggregation (or separation of mental processes) because the medium’s “inferior personality” (p. 219) manifested at a low intellectual level.

The séances he had with Palladino convinced Morselli that she could project a biopsychic force from her body, a force that could join with other forces coming from the other persons in the mediumistic circle (1908a, p. 350). This force could be imprinted with the “oniric or subconscious thought of the medium” (1908a, p. 243), which constituted the principle guiding telekinesis and shaping materializations. Morselli (1908b) believed that the subconscious thought and will of the medium directed the phenomena (p. 552). However, their uniformity and repetitive nature suggested to him that Palladino had “fixed ideas,” or delusional dominant ideas affecting both actions and thought. These ideas probably helped the production of the phenomena by her subconscious mind and also suggested hysteria (1908a, p. 251).

Morselli (1908a) admitted that all these psychological considerations did not really explain the so-called
biopsychic effects, that is, that an “unknown modality of energy . . . is exteriorized and projected into space” (p. 266). In his view the “constitutional morbidity” of mediums and their trances was not enough to explain physical phenomena nor veridical perceptions (1908b, p. 543). While Morselli believed that the subconscious mind was the domain of abnormal states such as ecstasy, hallucinations and the hypnotic trance, he was unsure if it was involved with the actual exteriorization of force. Physical phenomena, he noticed, could occur in different states of consciousness, such as the waking state or superficial or profound hypnotic states (1908a, p. 268). During superficial hypnotic states “Eusapia . . . says incoherent phrases, has muscular shudders, changes the tone of voice, has the fixed and glassy look of an hypnotized person . . .” (1908a, p. 272). During profound hypnotic states—a somewhat less developed version of Charcot’s lethargic stage of hysteria, visible materializations take place. The exteriorization of the force seemed to Morselli to be stronger the less the medium’s consciousness participated in the process. This led him to hypothesize that the trance inhibited the superior centers, leaving the inferior ones to act automatically with the force. This automatic action, Morselli (1908a) considered, was an inferior form of mental activity related to the “narrowing” of consciousness from the conscious mind (p. 322).

Further Speculations on Mediumship

Former Attorney General of Bordeaux, France, and physician Joseph Maxwell (1858–1938; see Evrard, 2009a) was influential in European psychical research (Maxwell, 1903/1905). Like many other “metapsychists” of his times, Maxwell had séances with Palladino and became convinced of the validity of the whole range of mental and physical mediumistic effects.

One of Maxwell’s contributions was his discussion of personifications, or the adoption of secondary personalities in mediumistic manifestations (Alvarado, 2011b). This included spirit controls, spirit communicators, and any manifestation that claimed to have a personality of its own. The phenomena, Maxwell said, “may personify God, the devil, angels, legendary personages, fairies, etc.” (Maxwell, 1903/1905, p. 64). These personifications could change with the composition of the mediumistic circle. On occasion, “sort of a collective consciousness is formed” (p. 65). Presumably coming from the sitters themselves, such a phenomenon “forms part of an . . . undeciphered chapter on the psychology of the crowds” (p. 65). According to Maxwell, this personification could appear with mediumistic communications as well as with physical phenomena such as raps indicating intelligence.

Maxwell saw in the mediums he studied an impressionability or nervous instability favorable for the phenomena. But Maxwell did not mean pathology such as that found in hysteria, neurasthenia, and other nervous affections. Another important researcher and theoretician was Swiss psychologist Théodore Flournoy (1854–1920). Flournoy wrote about the combination of subliminal and telepathic resources:

A medium personifies a dead person whom he has never known in a manner so admirable that it carries conviction to the sitters. They do not dream that perhaps one of them carries with him a group of memories which at the very moment that they are organized and flashed upward in the form of a composite portrait of the departed is telepathically reflected in the subconsciousness of the medium, as a living mirror, which immediately translates this image or imprint into words and gestures, portraying without a doubt a certain resemblance, but one in which the defunct has no share.

Combine, now, the facts of mental transmission with the products to which memory and subconscious imagination of the sitters can give birth, and you will understand what unforeseen complications may always be expected in spiritistic séances. So extraordinary do the revelations appear that it is very difficult to exclude the possibility that they are due to a play of action and reaction between the medium and the other persons present . . . (Flournoy, 1911, pp. 212–213)

Flournoy argued that the phenomena of psychical research could be explained by the subconscious powers of the living human mind. These powers, he said, manifested in certain individuals and “under certain psychophysiological conditions (secondary states of consciousness, somnambulism, etc.)” (Flournoy, 1911, p. 323). Flournoy, always open minded, said that this did not mean discarnate agency was not possible, but he emphasized that “before daring to give a verdict in favor of the discarnate, much more must be known as to the foundation and groundwork of our own constitution, with all its resources and endless possibilities . . . .” (Flournoy, 1911, p. 323)

The most scientifically eminent of the writers and researchers to be discussed in this section was the French physiologist Charles Richet (1850–1935). Richet (1883) experimented widely with hypnosis, and he witnessed and
induced many changes in personality using suggestion.

Like other individuals discussed in this paper, Richet accepted both mental and physical mediumship as genuine beyond dissociative explanations. Nonetheless, he was not a believer in spiritualism. To the contrary, he was a well-known skeptic on the issue of survival of bodily death in general, and of the question of spirit agency in mediumship in particular, although it is sometimes argued that he changed his view later in life (e.g., Alvarado, 2009).

Richet argued that in mediumship “a new personality is created by auto-suggestion” (Richet, 1923, p. 41). To him, mediums’ personifications, including spirit controls, could be explained by the workings of the unconscious mind. Referring to Piper, Richet wrote: “No doubt the personalities that present themselves . . . showed impressive traits of psychological individuality, and kept them distinctively in writing, voice, style, and thought” (p. 144; see also Richet, cited by Leaf, 1890, pp. 618–620). Many of these cases were seen by Richet to be an example of the “talents of the unconscious,” considered to be more creative than the conscious mind (p. 44).

Going beyond typical dissociative phenomena, Richet argued that in some mental mediums “metapsychic cognitions group themselves around the personality created by auto-suggestion” (p. 77). This assumed first the existence of telepathy and clairvoyance (or cryptesthesia, as Richet called it) in normal individuals, shown to occur both in real life as well as under experimental conditions, and secondly, an organization of these cognitions into a personality structure.

Another European psychical researcher, the Italian biologist William Mackenzie (1877–1970; see Melton, 2001, p. 952), published an interesting treatment of this topic in *Metapsichica Moderna*. Like previous writers, Mackenzie (1923) was clear in stating that in mediumship, the dissociative process presented in some cases phenomena beyond current scientific explanations (p. 162). Dissociation sometimes included a supernormal quality whose nature had not been elucidated but that allowed some phenomena to manifest “outside of the somatic system of the subject” (p. 167). This quality or element may be active in normal individuals while awake, in dreams and through intuitions. It may “assume great importance in some hypnotic states and in somnambulism; finally, it may be of maximum importance during the deep trance of the completely developed medium” (p. 168).

Regarding table tipping, Mackenzie referred to the “formation of a polypsic collectivity” (p. 183) formed by the sitters around the table. This psychic *quid* formed during a séance was not the simple addition of the subconscious of the medium and the sitters. Both were necessary to produce phenomena. “But the agent . . . is a *product* (and not an addition . . .) the ‘polypsic person’ . . . .” (p. 286). This was not simply the addition of separate elements but a new collective, a new entity from the dissociated aspects of the medium and the sitters.

Like Ochorowicz before him, Mackenzie theorized that the medium showed two types of dissociation. One was mental dissociation while the other was a new form of dissociation, a dissociation of matter taking place as a “function of a concomitant dissociation of the psyche” (p. 200). This physical dissociation consisted of a molecular and atomic decomposition of the medium’s body to create telekinetic and materialization phenomena.

Another psychical researcher who wrote about these ideas was the Frenchman René Sudre (1880–1968; see
Evrard, 2009b), who was also known as a science writer. In his textbook *Introduction à la Métapsychique Humaine*, Sudre (1926a) wrote about “prosopopesis,” or “brusk, spontaneous or provoked changes of psychological personality” (p. 85). This referred to the appearance of new personalities in mediumship, hypnotic states, possession cases, and in cases of double and multiple personality (see also Sudre, 1926b, 1946). The metapsychic type of prosopopesis was different from other types of dissociative personation phenomena in that sometimes it could be accompanied by phenomena such as “metagnomy” (clairvoyance) or materializations.

According to Sudre the association between metagnomy and prosopopesis made sense because both occurred during dissociation. Metagnomy, he thought, was common during trance and hypnosis, whereas prosopopesis was favored by the reduction and retraction of the waking consciousness. Although the combination of metagnomy and the dramatization of an external personality had convinced many that spirits of deceased individuals communicated through a medium, Sudre believed that metagnomy and prosopopesis, although perhaps sharing a common psychophysiological process, were different, independent functions.

As Sudre wrote, there are phenomena of prosopopesis “free of all supernormal element . . . where the subject adopts the speech and manner of a person whom he knows or fancies and by whom he claims to be invaded” (Sudre, 1926b, p. 129), and metapsychic phenomena without personality changes. In his words:

> Between these two extremes come the spiritistic phenomena which proceed from an apparent synthesis of the two functions. But we can always separate them, as does the chemist when he decomposes water into oxygen and hydrogen. We thus show that the medium creates in himself a new personality which is like the character personated in proportion as his faculty of clairvoyance allowed him to get information. If experience did not demonstrate that the two functions exist independently, if there were only spiritistic phenomena, we should not be entitled to make this distinction. (Sudre, 1926b, p. 130)

Physical phenomena were also related to personation and to dissociation. Sudre argued that a force exteriorized from the body was guided by the medium’s mind. In his words, “the subject’s intelligence, both conscious and unconscious is related to one of the dissociated layers of his personality” (Sudre, 1926a, p. 290).

In addition to some of the above mentioned individuals, many others accepted that some mediumistic phenomena could be explained by the supernormal powers of the living, while at the same time defending strongly the influence of discarnate spirits on mediumship (e.g., Bozzano, 1926; Delanne, 1902; Hyslop, 1901; Thomas, 1922).

**Concluding Remarks**

As argued before (Alvarado, 2002; Le Maléfan, 1999; Méheust, 1999b) ideas of the supernormal as regards dissociation and the subconscious were not integrated into the psychology and psychiatry of the times discussed in this paper. Although most medical men held a closed model of the mind (and of dissociation) in which the phenomena were explained mostly by internal resources and a few external influences such as suggestion, few accepted a more open model of mind, such as the one some psychical researchers upheld based on powers that extend sensory and motor capacities beyond the confines of the body. Nonetheless, and as seen in the writings of some such as James (1890a), these psychic or supernormal concepts were part of the same general interest in understanding the mind and its myriad of layers as the more accepted ideas of individuals such as Janet (1889). This situation has been documented by Ellenberger (1970) and by later historians (Crabtree, 1993; Gauld, 1992; Le Maléfan, 1999; Méheust, 1999b; Plas, 2000).

Interestingly, these ideas about the powers or capabilities of the subconscious mind were also connected in some cases to pathology. This was not only the case with those, like Janet (1889), who reduced everything to intrapsychic concepts, but also with those like Lombroso (1909) and Morselli (1908), who admitted the existence of the supernormal as a process related to pathologies such as hysteria. But most of the persons discussed here did not write about pathology.

The topic discussed here is also a reminder that the functions of the subconscious listed by Ellenberger and mentioned at the beginning of the article are incomplete without consideration of the psychical research perspective. In the period discussed here, psychical researchers considered that the functions of the subconscious went beyond memory, pathology, creativity, and imagination. In the case of mediumship, psychical researchers extended current ideas about dissociation (in this case trance and personation) by adding the supernormal to the equation.
We should keep in mind that for the period discussed here “supernormal” meant both the extended abilities of the living, such as telekinesis and telepathy, as well as discarnate agency. In other words, the trances of mediums such as Piper and Palladino were believed to facilitate both extraordinary human faculties and the workings of spirits. These ideas, I must add, were involved in controversy that I have not discussed in this paper. This included the anti-spiritualistic writings of some psychical researchers (Richet, 1923; Sudre, 1926a), the criticisms that believers in spirit action presented about ideas of secondary consciousness, dissociation, and telepathy (e.g., Delanne, 1902; Noel, 1885), and the writings of those skeptical of the supernormal in its totality (e.g., Janet, 1889; Jastrow, 1906).

It is my hope that the material discussed in this paper will remind current students of mediumship of aspects of a past forgotten by many. Furthermore, I hope that my writings and those of others will influence the general historiography of psychiatry and psychology, a trend clear in the work of a few writers (e.g., Crabtree, 1993; Plas, 2000). Currently most of this work refers to the “closed” model of the mind and of dissociation represented by Janet and others. But to limit historical analysis in this way produces an incomplete picture of the past, the details of which are ignored or dismissed by many historians as well as by psychiatrists and psychologists. Regardless of the many subjective aspects of the study of history, attempts to study ideas about the mind should seek to represent the past in its own terms, a history which most certainly includes ideas such as those discussed in this paper.

References

Mediumship, Psychical Research, Dissociation, and Powers of the Subconscious Mind

Psychical Research, 6, 651–659.


Acknowledgements

I wish to thank the Society for Psychical Research for a grant that allowed me to prepare this paper. Thanks are also due to Debra H. Weiner for useful editorial suggestions.

Abstracts in Other Languages

German

MEDIALITÄT, PSYCHISCHE FORSCHUNG, DISSOZIATION UND DIE KRÄFTE DER UNTERBEWUSSTEN PSYCHE


Spanish

MEDIUMNIDAD, INVESTIGACIÓN PSÍQUICA, DISOCIACIÓN, Y PODERES DE LA MENTE SUBCONSCIENTE

RESUMEN: Desde el siglo 19 muchos psiquiatras y psicólogos han considerado que la mediumnidad está relacionada con la mente subconsciente y procesos disociativos producidos principalmente por los procesos convencionales internos de la mente del médium. Sin embargo algunos psicólogos e investigadores psíquicos activos entre las últimas décadas del siglo 19 y la década de 1920 expresaron una opinión diferente. Individuos como Théodore Flournoy, Cesare Lombroso, Enrico Morselli, Frederic W. H. Myers, Julian Ochorowicz, Charles Richet, Eleanor Sidgwick y Eduard von Hartmann argumentaron que algunos médiums combinan la disociación con fenómenos sobrenaturales tales como el conocimiento adquirido sin el uso de los sentidos y la producción de efectos físicos aparentemente más allá de las capacidades corporales normales. Dependiendo del teórico en cuestión, otras cuestiones como la patología y agencias desencarnadas también formaban parte de las discusiones. Lo supernormal nunca fue aceptado por la ciencia en general y en la actualidad rara vez se menciona en la literatura sobre disociación, pero las ideas relacionadas con lo supernormal fueron parte de esta literatura. La historia completa de la disociación y la mente subconsciente deben considerar este cuerpo de trabajo.
MÉDIUMNITE, RECHERCHE PSYCHIQUE, DISSOCIATION, ET LES POUVOIRS DE L’ESPRIT SUBCONSCIENT

RESUME : Depuis le XIXe siècle, de nombreux psychiatres et psychologues ont considéré la médiurnité comme étant liée à l’esprit subconscient et aux mécanismes dissociatifs produits surtout par des processus internes conventionnels dans l’esprit du médium. Toutefois, certains psychologues et parapsychologues du tournant du XXe siècle ont exprimé une vue différente. Des individus tels que Théodore Flournoy, Cesare Lombroso, Enrico Morselli, Frederic W.H. Myers, Julian Ochorowicz, Charles Richet, Eleanor Sidgwick et Eduard von Hartmann, ont affirmé que certains médiums combinaient la dissociation avec des phénomènes supernormaux, tels que la connaissance acquise sans l’emploi des sens, et la production d’effets physiques semblant déborder les capacités somatiques normales. Dépendamment du théoricien, d’autres questions relatives à la pathologie ou aux agents désincarnés furent également inclues dans les discussions. Le supernormal ne fut jamais accepté par la science dans son ensemble et, aujourd’hui, il est rarement mentionné dans la littérature sur la dissociation. Mais les idées associées au supernormal constituent une part de la littérature sur ce sujet. Une histoire complète de la dissociation et de l’esprit subconscient ne peut faire l’impasse sur ce corpus.
ABSTRACT: This online dream precognition study examined variables, both psychological and parapsychological, that have been proposed to contribute to precognitive dream experiences. 50 participants each contributed 4 trials, where the task was to dream about a video clip that they would later view. Independent judges were used to score the correspondence between dreams and the target pool. No support was found for the hypothesis that individuals who are intolerant of ambiguity would report greater correspondence between their dreams and subsequently viewed target video clips. A relationship was found between the participants’ prior confidence that their dreams would relate to the future target and actual perceived similarity between the target and dreams; however, there was no relationship between perceived similarity and judges’ actual hit rates or similarity ratings. The test of the precognition hypothesis obtained above-chance scoring (32% hit rate) on the planned direct hits measure. Obvious methodological artifacts are ruled out, and the discussion concludes with an exploration of whether the judges’ ratings also support the dream precognition hypothesis.

Keywords: precognitive dreaming, precognition, precognitive dream experiences, ambiguity tolerance

Surveys of the general population show that reports of psi-related experiences such as apparent clairvoyance, telepathy, and precognition are common throughout the world. For example, a 1987 survey published by the University of Chicago’s National Opinion Research Center canvassed nearly 1,500 adult Americans, of whom 67% claimed psi-related experiences (Greeley, 1987). Precognition—seemingly knowing about an event that has yet to take place—was reported by approximately one third of respondents in a recent survey of 1,000 Britons (Pechey & Halligan, 2012).

Dreams seem to play a particularly important role in precognitive experiences. A review of the various surveys of spontaneous GESP experiences concludes that, if only precognitive cases are considered, around 60% involve dreams, with a further 10% involving “borderland” states (Van de Castle, 1977). Therefore, the vast majority of spontaneous precognitive experiences involve dreams or sleep-related states. Death is a predominant theme in precognitive dreams, followed by accident and injury; percipients are predominantly female (e.g., Green, 1960; Saltmarsh, 1934), although reporting bias may account for both of these trends.

When considering possible explanations for spontaneous paranormal experiences, researchers often either consider a paranormal interpretation, or one of several possible psychological explanations, although these are not mutually exclusive categories. Researchers tend to turn to controlled laboratory settings to test the psi hypothesis. Only a minority of laboratory dream ESP studies have investigated precognition, which is perhaps odd given the prevalence with which spontaneous dream precognition experiences are reported. Controlled laboratory studies of dream ESP took off from 1962, after psychiatrist Montague Ullman established a dream laboratory at the Maimonides Medical Center in Brooklyn, New York (Krippner, 1993; Ullman et al., 1973, 1989). Thirteen formal dream ESP studies (11 telepathy, 2 precognition) were conducted at the lab before it closed in 1978, the majority of which obtained medium to large positive effect sizes (Sherwood & Roe, 2003). A review of the 21 post-Maimonides dream ESP studies identified that, for the majority of them, the research environment had moved from the relatively expensive and time-consuming sleep laboratory to participants’ own homes (Sherwood & Roe, 2003). The studies had a modest combined effect size (r = .14) —significantly less than for the Maimonides studies, but still regarded as “successful” by Sherwood and Roe, who expressed the hope that dream ESP research would be “re-awakened.”

1An earlier version of this paper was presented at the 2011 Conference of the Society for Psychical Research in Edinburgh, September 3–5.
Turning to possible psychological factors underlying paranormal experiences, Blackmore and Moore (1994) proposed that paranormal believers and disbelievers might have different cognitive styles. They tested this idea by presenting participants with ambiguous pictures and found that believers guessed the identity of the picture earlier than disbelievers, though the believers were more often incorrect in these guesses. In this study, therefore, believers tended to rapidly evaluate the ambiguous stimuli and, compared to disbelievers, set a lower criterion for identifying these patterns. This propensity seems related to intolerance of ambiguity, which is conceptualised as a form of premature closure achieved through a tendency to resort to clear-cut solutions in ambiguous situations (Frenkel-Brunswick, 1949). Houran and Williams (1998) explored the relation between ambiguity tolerance and specific paranormal experiences using Kumar, Pekala, and Gallagher’s (1994) Anomalous Experiences Inventory and MacDonald’s (1970) Ambiguity Tolerance scale. They reported that there was a small but positive correlation between experiences involving internal or physiological experiences, such as precognitive dreams, visual apparitions, and out-of-body experiences, and tolerance for ambiguity. This finding seems to be inconsistent with Blackmore and Moore’s (1994) conclusion, leading Houran and Williams (1998) to suggest that variability in the measures used across different studies may contribute to the equivocal association between ambiguity tolerance and paranormal beliefs and experiences. They called for further work on this question.

The present paper describes work that answers colleagues’ calls and build upon both of these lines of investigation. Researchers are now able to present stimuli and collect data rapidly from participants online, something that is particularly practical when investigating dream precognition. Sleeping in their own homes, the participants kept dream diaries and used a website to complete questionnaires and submit dream summaries and ratings at times that were convenient to them. Email was used to coordinate and communicate with participants; target feedback was rapidly given via YouTube. Participants were asked to complete questionnaire measures concerning their precognitive dream experience, dream recall, and tolerance of ambiguity. Their task was then to dream about a target video clip that would subsequently be sent to them. They submitted a weekly dream summary that was rated for similarity with randomly chosen target pools by independent blind judges. Independent judges were used because if participants were to be judges, they would see all target possibilities; therefore, their dreams could in theory precognise one of the decoy targets. Having participants only view their designated target video was, one felt, a way to “focus” any precognition on the target. After the judges had made their ratings, a target video clip was randomly selected and sent to participants, who were not informed of the judges’ ratings. After viewing the target clip, participants were asked to rate it for similarity with their dreams.

Two hypotheses were proposed. First, to test the idea that dreams can contain unpredictable information about future events, it was hypothesised that there would be significantly more direct hits than chance, based on the independent judges’ ranks of the target and three decoy clips. Second, to explore the psychological factor of ambiguity tolerance, which has been proposed to contribute to precognitive dream experiences (Blackmore & Moore, 1994), it was hypothesized that there would be a negative correlation between ambiguity tolerance and the participants’ target clip similarity ratings.

**Method**

The study was approved by the University of Edinburgh’s Psychology Department ethics panel.

**Participants**

Participants were recruited through posts on Twitter, by email sent to former participants of an online parapsychology course led by CW, through the KPU website, from amongst acquaintances of the authors, and by word of mouth. Individuals were invited to volunteer if they were interested in their dreams (precognitive or otherwise) and able to recall their dreams. The co-experimenter (MV) sent participants detailed information on the study prepared by CW. Volunteers received no financial reward for participating in the study.

**Independent Judges**

Two individuals who had an interest in parapsychology (a psychology PhD student and a psychology undergraduate student) acted as independent judges. They were each paid for their work as judges. Because they
had no previous experience in judging, they were given guidelines on free-response ESP judging (Delanoy, Morris, & Watt, 2004).

Materials

Participants provided three types of information via online questionnaires: demographics, beliefs, and ambiguity tolerance measures; a dream summary and confidence ratings; and similarity ratings.

Questionnaire 1: Initial questionnaire. The initial questionnaire consisted of two parts. In the first, demographic, part the participants indicated:

1. Age and sex.
2. Frequency of dream recall: “How often have you recalled your dreams recently (in the past several months)?” Response options for the latter were “never,” “less than once a month,” “about once a month,” “two or three times a month,” “about once a week,” “several times a week” and “almost every morning.” This scale was developed by Schredl (2004), who reports a high test-retest reliability over approximately 70 days, $r(196) = .85$. Scores could range from 0 to 6, with higher scores denoting more frequent dream recall.
3. Belief in precognitive dreaming: “Do you believe that some individuals have dreams that predict future events, and that are not just coincidence?” Bender’s (1966) five criteria for judging a dream as precognitive were provided to participants, along with parenthetical explanations where appropriate, in order to ensure conceptual clarity: (a) the dream must be recounted or recorded before its fulfilment (e.g., was it written down or described to another person before it “came true”); (b) the dream must include enough details to render chance coincidence unlikely; (c) the possibility of inference from actual knowledge must be excluded (i.e., the dream must refer to an unexpected or unpredictable event); (d) self-fulfilling prophecies must be excluded (i.e., you could not make the dream “come true” through your own actions after the dream); and (e) telepathic influences should not be able to explain the occurrence of the precognitive dream (i.e., no one else could know the information in the dream at the time that you had the dream). The response options were “yes,” “no,” and “unsure.”
4. Frequency of precognitive dream experience: “Based on the five criteria above, please indicate approximately how often you have had a precognitive dream over the last few years.” Response options were “never,” “less than once a year,” “about once a year,” “about once in six months,” “about once a month,” and “about once a week.” Scores could range from 0 to 5, with higher scores denoting more frequent experiences.

The second part of the initial questionnaire consisted of the Revised Scale for Ambiguity Tolerance (AT-20; MacDonald, 1970). It is a 20-item first-person statement inventory. A sample item is: “A problem has little attraction for me if I don’t think it has a solution,” with “true” and “false” as response options. The scale includes 15 reverse-scored items. Scores on this scale can in theory range from 0 to 20, with higher scores indicating higher tolerance for ambiguity. MacDonald reports high internal consistency of .86, adequate test-retest reliability ($r = .73$), and satisfactory stability over six months ($r = .63$). However, in the present sample the internal consistency of the scale was somewhat lower ($\alpha = .66$). After the exclusion of three negatively or weakly correlated items (items 4, 5 and 7), the internal consistency rose to $\alpha = .72$. This trimmed version of the scale was used for further analysis.

At the conclusion of the initial questionnaire, participants were instructed to take note of their dreams over the following 5 mornings. They were informed that after 5 days they would be sent a link to a questionnaire asking for an anonymous summary of their week’s dreams. They were reminded that: “after we have received your dream summary you will be sent a ‘target’ video clip to view. Every night, before you go to sleep, please take a few moments just to gently remind yourself that your dreams during the night will be linked to the target clip you are going to watch after we have received your dream summary.”

Questionnaire 2: Dream summary form. This form, which participants were provided with after 5 days, consisted of three items:

1. Participants’ weekly dream report: “Please type in the space below an anonymous summary (max 300
words) of your remembered dreams over the past 5 days. Include not only descriptions of main content and themes that emerged in your dreams, but details such as emotional tone and the impact of the dream. You do not need to write anything that you would find embarrassing or that would make you uncomfortable to write. Please don’t provide any personally identifying details.”

2. Confidence rating: “Please rate how confident you are that your dreams over the past 5 days will relate to the target video clip that you will shortly be sent.” Response options were “not at all confident,” “not very confident,” “somewhat confident,” “very confident,” and “completely confident.” Scores could range from 1–5, with higher ratings denoting higher confidence. Participants were also asked to explain why they chose their particular confidence rating.

**Questionnaire 3: Similarity rating form.** After participants had been sent a link to their target video, they were asked to “indicate how much similarity you feel there is between your submitted dream summary for this week and your target video clip for this week. Please bear in mind not just dream content, but associated themes and emotions.” Participants typed in a number between 1 and 100, where 1 = no similarity and 100 = complete similarity.

**Target Pool**

The stimulus pool consisted of 68 short (around 1 min) video clips divided into 17 target pools of four orthogonal videos, each uploaded to YouTube. The target clips were digitized from a pool used in KPU ganzfeld-ESP research that had obtained positive psi results (e.g., Dalton, 1997; Morris, Dalton, Delanoy, & Watt, 1995). The pool included scenes from films, nature documentaries, and music videos. There had originally been 18 target pools (i.e., 72 video clips), but one was withdrawn prior to the commencement of the study after a copyright query was raised by YouTube.

**Random Number Generator**

For random selection of the target pools and targets, an RNG function from the website RANDOM.ORG was used. It generates numbers based on atmospheric noise and is therefore a true random source. It is more appropriate for a precognition study than a pseudo-random source, because it rules out the possibility of clairvoyance.

**Procedure**

The initial questionnaire, along with the two forms, was published online using the Google Forms service. Participants could therefore complete the online questionnaire and forms after being sent the appropriate URL. The target pool was uploaded to a YouTube channel and marked as unlisted, so that targets were accessible only via a particular URL.

The experimenters and judges took part in pilot trials to refine and test the protocol. The results of the pilot trials are not included in this report.

The study consisted of 200 trials (preplanned as four trials each from 50 participants). For security reasons, the target for any one trial was randomly selected and sent to participants only after the independent blind judge had submitted his rating of the four randomly chosen target pool videos against the dream summary for that trial. Therefore, there could be no leakage of target information, either from the randomiser to the judges or from the participants to the judges.

Each participant was assigned by MV to one of the two judges and was sent a hyperlink to the initial questionnaire. Participants had no contact with the judges, nor were they aware of the judges’ identity; likewise, the judges were unaware of the participants’ identities. After completing and submitting the initial questionnaire, the participants were informed that their 5-day dream collection period had commenced. On the fifth night they were sent an email informing them that the dream collection period was about to end and that their first dream summary was due the next morning. They were also sent a hyperlink to the dream summary form. Upon receiving the dream summary from a participant, MV randomly selected a target pool for that participant (1 of the 17) and sent the anonymised dream report along with the URLs of the target pool videos to the judge. Within each target pool, the number of the clip determined the position in which its URL would be presented to the independent judges. So, for
target pool 1, clip 1-1 would be first in the list of four URLs, clip 1-2 second, 1-3 third, and 1-4 fourth. Judges could (and did) review the four clips in whatever order they chose, and could (and did) view the clips more than once during the judging process for any particular trial.

The judges were instructed to provide a percentage rating of the similarity between each of the four videos in a given target pool and the contents of the dream summary, as well as a ranking of the videos based on these ratings (rank 1 = greatest similarity, rank 4 = least similarity). No tied ratings were permitted, and a hit was defined as a rank of 1 corresponding to the designated target. They subsequently emailed their judgements to MV in an attached file. MV then, without viewing the judgements, randomly selected a target video from the given pool and sent its URL to the participant via email. The participants were also instructed to follow a hyperlink to the dream similarity rating form upon viewing the target videos.

Two to 3 days after receiving the participants’ similarity rating, MV informed the participants by email that the second dream collection period was commencing and the procedure repeated itself. Altogether, for each participant, four trials were conducted over approximately a 4-week period. Throughout the study, the participants were thanked for their involvement and indirectly encouraged to continue. Participants who failed to submit either of the forms were sent a gentle reminder to do so.

At the conclusion of the study, as soon as data had been checked and analyses had been completed, participants were sent a short summary of the overall study results. They were not informed of the outcome of their judge’s ratings while the study was underway.

Results

A total of 99 volunteers were sent the link to the initial questionnaire. Twenty-two of them did not return a completed questionnaire so did not proceed with the study. Recruitment continued until 50 participants (20 males, 30 females; mean age 42.8, range 21–82 years, \(SD = 14.41\)) had completed four trials each. Twenty-one others dropped out of the study before completing four trials; 6 completed four trials after the pre-planned \(N\) of 50 participants had been reached. Data for these 27 participants are not included in this report. Prior to analysis, the scoring of the questionnaire measures was independently checked for possible errors, as was the recording of the study’s psi results (both the judges’ ratings and rankings, and the subsequent assignment of each trial as a hit or miss).

Descriptive Statistics

Table 1 shows \(M, SD, N\) and range of scores on the principal questionnaire measures.

<table>
<thead>
<tr>
<th></th>
<th>(M)</th>
<th>(SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>42.82 years</td>
<td>14.41</td>
<td>21–82</td>
</tr>
<tr>
<td>Dream Recall Frequency (0–6)</td>
<td>4.88</td>
<td>1.10</td>
<td>1–6</td>
</tr>
<tr>
<td>Precognitive Dreams Belief (0–2)</td>
<td>1.58</td>
<td>0.64</td>
<td>0–2</td>
</tr>
<tr>
<td>Prior Precog. Dream Experience (0–5)</td>
<td>1.70</td>
<td>1.53</td>
<td>0–5</td>
</tr>
<tr>
<td>Ambiguity Tolerance (0–20)</td>
<td>11.02</td>
<td>3.57</td>
<td>3–17</td>
</tr>
<tr>
<td>Mean Confidence Rating (1–5)</td>
<td>2.04</td>
<td>0.74</td>
<td>1–3.75</td>
</tr>
<tr>
<td>Mean Similarity Rating (1–100)</td>
<td>15.15</td>
<td>15.95</td>
<td>0.25–60</td>
</tr>
</tbody>
</table>

*Note. \(N = 50\) in all cases.*
**Dream recall.** A large proportion of participants (50%) reported that they recalled their dreams several times a week, followed by almost every morning (28%), about once a week (10%), two or three times a month (8%), once a month (2%), and less than once a month (2%). As participants needed to be able to remember their dreams to successfully participate in the study, it was reassuring that the majority (88%) remembered their dreams at least once per week.

**Precognitive dream belief.** Having been asked to use Bender’s (1966) criteria for evidentiality, most participants (66%) expressed a belief that individuals could have precognitive dreams, 26% were unsure, and 8% did not believe in precognitive dreams. Thus, the sample was skewed towards individuals believing in precognitive dreams. Because there were so few disbelievers in the sample (4 out of 50), no attempt was made to compare disbelievers and believers on the other study measures.

**Precognitive dream experience.** Twenty-eight percent of participants indicated that they had never had a precognitive dream experience (again, as defined by Bender’s criteria), 26% less than once a year, 14% about once a year, 16% about once in 6 months, 12% about once per month, and 4% about once a week. Thus, the majority of participants (72%) reported having at least one prior precognitive dream experience that would be considered evidential.

**Confidence ratings.** Participants were not particularly confident that their dream summaries would relate to the future target video (mean rating 2.04 on a scale from 1–5). Participants who gave low confidence ratings reported that they did so either because they did not believe in dream precognition, or because their own precognitive dreams were more personal and they did not think their dream would relate to a randomly selected target video.

**Similarity ratings.** Following feedback about the target video identity, participants’ similarity ratings suggested that they saw little similarity between their dream summaries and the target videos (mean rating 15.15 on a 1–100 scale).

**Main Analyses**

**Dream precognition (Hypothesis 1).** Direct hits analysis was planned (rather than, for instance, sum-of-ranks, binary hits where rankings or ratings in the top half = binary hit and in the bottom half = binary miss, or z-score based on judges’ ratings), firstly because Child’s (1985) meta-analysis of Maimonides dream-ESP studies used direct-hits outcome measure to allow comparison between studies, and secondly because participants only viewed the target videos so it was predicted that any precognitive dream content would focus on these videos. As it turns out, the decision to base analyses on hits rather than ratings did not disadvantage the psi hypothesis: Judges’ ratings of the targets and decoys did not show elevated ratings for the target video clips relative to decoy clips, Mann-Whitney $U = 56073.5, p = .16$, two-tailed. Sixty-four hits were obtained out of 200 trials, giving a 32% hit rate. Using an exact binomial test, this result is significant, $z = 2.21, p = .015$, one-tailed, ES ($z/N^{1/2}$) = 0.16. Thus, Hypothesis 1 was supported.

**Ambiguity tolerance and similarity ratings (Hypothesis 2).** The Mean AT score was 11.02, range = 3–17, $SD = 3.57$. Contrary to expectation, there was no significant relationship between AT and participants’ mean similarity ratings; indeed, the correlation was in the direction opposite to that predicted, $r(48) = .158, p = .27$, two-tailed. Therefore, Hypothesis 2 was not supported.

**Exploratory Analyses**

**Independent judges.** There was little difference between the hit rates obtained by the two judges (Judge 1, 26 hits out of 84 trials, 31.0%; Judge 2, 38 hits out of 116 trials, 32.8%), as confirmed by an independent groups $t$ test comparing the judges’ mean numbers of hits per participant, $t(48) = 0.32, p = .75$, two-tailed.

**Prior dream recall, precognitive experience, confidence and similarity ratings.** Older participants tended to report having had more precognitive experiences than younger participants, $r(48) = .29, p = .04$, two-tailed. Also, as one might expect, there was a significant tendency for participants reporting greater numbers of prior precognitive experiences to give higher ratings of confidence that their dream reports would contain material relating to the future target video, $r(48) = .32, p = .02$, two-tailed. However, as these two relationships were not predicted, and as several correlations were calculated for the psychological variables (see Table 2), it would be wise to regard them as only tentative, in need of replication.
Table 2
Matrix of Spearman Correlations Between Study Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>-.19</td>
<td>.29*</td>
<td>-.06</td>
<td>-.12</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td>Dream Recall</td>
<td>-.19</td>
<td>.16</td>
<td>-.04</td>
<td>.16</td>
<td>-.06</td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>Precognitive Dream Experience</td>
<td>.29*</td>
<td>.16</td>
<td>-.07</td>
<td>-.01</td>
<td>.32*</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Ambiguity Tolerance</td>
<td>-.06</td>
<td>-.04</td>
<td>-.07</td>
<td>.19</td>
<td>-.04</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Total Hits</td>
<td>-.12</td>
<td>.16</td>
<td>-.01</td>
<td>.19</td>
<td>-.14</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Mean Confidence</td>
<td>.06</td>
<td>-.06</td>
<td>.32*</td>
<td>-.04</td>
<td>-.14</td>
<td>.41**</td>
<td></td>
</tr>
<tr>
<td>Mean Similarity</td>
<td>.01</td>
<td>-.07</td>
<td>.07</td>
<td>.16</td>
<td>.03</td>
<td>.41**</td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 50 in all cases.
*p < .05. **p < .01. All two-tailed

After receiving feedback of the target video’s identity and giving it a rating for similarity to their previously submitted dream report, participants who had previously given higher confidence ratings tended also to give higher similarity ratings, \( r_s(48) = .41, p = .003 \), two-tailed. However, there was little relationship between these ratings and actual psi performance: The correlation between confidence ratings and hit rate was slightly negative \( r_s(48) = -.14, p = .34 \), two-tailed, and there was no correlation between hit rate and similarity ratings \( r_s(48) = .03, p = .86 \), two-tailed.

Self-reported prior dream recall did not significantly correlate with precognitive dream experience, confidence ratings, similarity ratings, or hit rate. Ambiguity Tolerance also did not correlate with any of these variables. Table 2 gives the full matrix of correlations for the variables reported here.

I also explored whether the participants may have given higher similarity ratings to the target clips than the judges, perhaps because the participants were better able to recognise their own dream content in the targets than the judges were. Although participants’ mean ratings were slightly higher than those of the judges, this was primarily attributable to a small number of “outlier” similarity ratings above the mid-point of the scale. The vast majority of ratings by judges and participants were strongly skewed towards the bottom end of the 100-point scale (judges’ median rating 4.5, \( SD = 12.35 \); participants’ median rating = 5.0, \( SD = 21.67 \)); a weak but significant correlation was also found between judges’ and participants’ similarity ratings, \( r_s(198) = .14, p = .04 \), two-tailed.

Discussion

The majority of individuals who took part in this study reported that they believed in precognitive dreaming, had experienced an evidential (according to Bender’s criteria) precognitive dream personally at least once in their lifetime, and were able to recall their dreams at least once per week. So, on the face of it, this sample would seem to be appropriate for a dream precognition study.

The study hypotheses explored two types of explanation for precognitive dream experiences: psychological and paranormal. For the first type, it has been proposed that individuals who are low in tolerance for ambiguity (AT) will be most likely to report that they have spontaneous precognitive dream experiences. The rationale for this prediction, from Blackmore and Moore (1994), is that these individuals will be more inclined to prematurely judge that there are “matches” between their dreams and subsequent events. This hypothesis was tested by correlating AT with the similarity ratings that participants gave to their dream summaries, having received feedback about the target identity. However, the data did not support the prediction of a negative correlation between AT and similarity ratings. A possible explanation of this null finding, other than the nonexistence of the hypothesised relationship,
The Journal of Parapsychology

concerns the low internal consistency of the ambiguity tolerance scale. Although the consistency reported by the authors of the scale is good, applied to the present sample the scale’s internal consistency index (Cronbach’s α) did not reach the conventionally acceptable value of .8, even after the exclusion of negatively and weakly correlated items. The present study’s finding is more consistent with Houran and Williams (1998), who reported a small but positive correlation between AT and paranormal experiences, including precognitive dreams; however, they do not report internal consistency for the AT-20. Lange, Schredl, and Houran (2000) have argued that there is some complexity to the relationship between ambiguity tolerance and precognitive dream experience, and they suggest that a nonlinear model may better describe the relationship. Such a model might therefore help to account for the mixed findings so far.

The study results did suggest the operation of psychological mechanisms that can lead to increased subjective experience of precognitive dreams. Participants who had higher confidence tended to report greater levels of similarity between their dreams and the target video, although perceived similarity was not associated with a higher hit rate or actual similarity ratings. So, prior confidence appears to be associated with perceived correspondences between dreams and subsequent events. Some previous research has suggested that frequency of dream recall is a factor likely to create more opportunities for correspondences between dreams and subsequent events. Some previous research has suggested that frequency of dream recall is a factor likely to create more opportunities for correspondences between dreams and subsequent events to be noticed (e.g., Lange, Schredl, & Houran, 2000). Contrary to this suggestion, the present study found only a weak positive correlation between reported dream recall and prior precognitive experience that did not reach statistical significance. One referee of this paper commented that there may be difficulties in interpreting participants’ responses to the dream recall measure (developed by Schredl, 2004) due to the nature of the question, which required retrospective reflection and self-report from participants (rather than, for instance, having them keep a diary and then count how often they remembered their dreams). However, this kind of self-report measure is common in psychological research despite the response bias that can accompany any such measure. Furthermore, Schredl reports high test-retest reliability, which indicates consistency in responses.

I also conducted a controlled test of the hypothesis that individuals’ dreams can contain information about unpredictable future events, in other words, that some form of anomalous cognition can occur. Independent judges rated each participant’s dream summary for similarity to the contents of a randomly selected pool of four video clips, and then one of these clips was randomly selected as the target and sent to the participant for feedback. Judges gave the highest similarity ratings to the future target clip significantly more often than would be expected by chance, thus supporting the hypothesis. An above-chance hit rate provides evidence for a psi process only if plausible alternative explanations can be ruled out, so these alternative explanations will now be considered.

Consideration of Alternative Explanations for the Significant Hit Rate

1. Judges were deliberately or unconsciously biased by the experimenter’s knowledge of the selected target. This explanation does not apply because the experimenter did not select the target until after the judges’ ratings were made.

2. The experimenter’s target selection was biased by his knowledge of the judges’ ratings. The experimenter did not view the judges’ ratings prior to target selection. Furthermore, target selection was done using an online random number generator, which would not under normal circumstances be influenced by the experimenter.

3. Participants leaked information about the target identity to the judges, for example, using online social networking sites. The judges did not know the identity of the participants and, even if they did, the judging was completed before participants were given feedback about the target identity.

4. Participants’ dream summaries contained cues as to previous weeks’ targets that may have leaked information to the judges about the target identity. The judges rated each trial on the day that the dream summary was received, so the judging was done in real time. This means that dream summaries could contain only information about previous targets that had already been judged. This information would not be useful for the present trial being judged.

5. The coordinating experimenter cheated. The records for each trial were independently checked and verified after the study was concluded. For cheating to apply, one therefore has to adopt an unfalsifiable conspiracy theory, including fraud by the judges and the principal investigator.
These five points demonstrate that the study design precludes the most obvious methodological flaws that might lead to spuriously significant results. Finally, one must consider whether this study’s significant hitting may reasonably be attributed to precognitive dreaming on the part of the study participants, or whether some as yet unexplained alternative form of psi or undetected methodological artifact may be at work.

Is a Precognitive Dreaming Interpretation Supported by the Study Data?

The precognitive dreaming hypothesis requires that judges detect a greater degree of similarity between the participants’ dreams and the designated target video clips, compared to decoy clips. If this were the case, then one would expect to see three things in the study data. First, one would expect the target clip to be given the highest ranking to a greater than chance extent. The study design prespecified direct hits (based on ranks) as the outcome measure, and a significant outcome was indeed found. This supports the precognitive dreaming hypothesis. Second, one would expect judges’ ratings for clips designated as the target to be greater than for clips designated as decoys. This would indicate that judges detected greater similarity between dreams and targets than between dreams and decoys. Third, there should be a difference in target versus decoy ratings for hit trials compared to miss trials, because in the hit trials judges presumably select the target because it is more similar to the dream mentation than the decoy clips. Further exploratory analyses of judges’ ratings that address the latter two questions will now be presented.

Analysis of judges’ ratings. The Results section provides a justification for the decision to analyse ranks rather than rating scores (1–100). It also shows that there was no significant difference between the ratings of targets and decoys. However, it could be argued in line with the psi hypothesis that this result could be expected. Even if there was indeed a communication anomaly, there is no reason to expect all targets to be rated as more similar than decoys; only hits should be particularly salient to judges, due to there being a noticeable similarity between the participant’s dream report and the target content. This reasoning, however, is problematic because, by definition, targets have the highest ratings in hit trials. Instead, one can explore the difference between the rating of the videos ranked 1 for hit trials and miss trials. If the saliency hypothesis is true, there should be a difference. However, in the present study, there was no such difference, \(U = 4127.5, p = .56\). Taking this one step further, it could also be argued that it is not the ratings per se that should differ. After all, it is quite possible that salience matters only in the context of other videos in the pool. If all of them are equally similar to the dream, then there is no particular salience for the target, and hence any hits are due to chance. On the other hand, the argument goes, hits obtained due to psi should be characterised by the target standing out from amongst the other video clips in the pool. If this is true, a comparison of the ratio of the rating for the video ranked 1 to the mean rating of the videos ranked 2–4 (Rank 1 / mean(Rank 2 - 4)) between hit and miss trials should reveal a difference. But again, the analysis did not yield significant results, \(U = 4318.5, p = .47\). The analyses reported above suggest that there was nothing qualitatively special about the hits compared to the misses.

In conclusion, on the preplanned direct hits measure, the study outcome is consistent with a precognitive dreaming interpretation. However on the exploratory analyses of ratings, there was no significant difference between the judges’ ratings of targets and decoys. Furthermore the ratings for targets that scored a hit were on average no more similar to the dream reports than the ratings for those that did not, whichever way one looks. These latter two observations are inconsistent with an interpretation in terms of precognitive dreaming, and may indicate the presence of nonpsi factors. However at best they can only tentatively qualify the planned outcome measure because (a) they are post hoc, and (b) they may simply indicate that ratings are a less reliable indicator of psi than rankings, for instance because, although they can offer a more fine-grained measure, they may also be more susceptible to extraneous “noisy” influences. As is always the case when a significant outcome is reported in a study using an original method to test the psi hypothesis, this study’s findings could be due to chance or an undetected artifact and should therefore be regarded as tentative pending replication.

References


Kumar, V. K., Pekala, R. J., & Gallagher, C. (1994). The Anomalous Experiences Inventory. Unpublished test, West Chester University, PA.


Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ, UK
Caroline.Watt@ed.ac.uk

**Acknowledgements**

I would like to thank Milan Valášek for his role in the logistics of the study as well as suggesting the rating analysis that is presented in the Discussion. I am grateful to the Perrott-Warrick Fund for making this research possible, and to the editor, statistical editor, Jim Kennedy, and anonymous referees for making helpful suggestions to improve the paper. Thanks to David Acunzo, Abigail Alfrey, Julia Cagle and Marius Leckelt for their help with the study, and to the study participants.
Precognitive Dreaming

Abstracts in Other Languages

German

PRÄKOGNITIVES TRÄUMEN: ZUR UNTERSUCHUNG ANOMALER KOGNITION UND PSYCHOLOGISCHER FAKTOREN


Spanish

SUEÑO PREMONITORIOS : INVESTIGACIÓN DE LA COGNICIÓN ANÓMALA Y LOS FACTORES PSICOLÓGICOS

RESUMEN: Este estudio en línea sobre sueños precognitivos examinó las variables, tanto psicológicas como parapsicológicas, que se han propuesto como contribuyentes a estos sueños. Cada uno de 50 participantes contribuyó cuatro pruebas, en donde la tarea era soñar con un corto de vídeo que verían más tarde. Jueces independientes evaluaron la correspondencia entre los sueños y la muestra de cortos. No se encontró apoyo a la hipótesis de que los individuos que no toleran ambigüedad reportarían mayor correspondencia entre sus sueños y los cortos vistos posteriormente. Hubo una relación entre la confianza previa de los participantes de que sus sueños se relacionarían con el objetivo futuro y la similitud real percibida entre el objetivo y los sueños; sin embargo no hubo relación entre la similitud percibida y la tasa de éxito o clasificaciones de similitud de los jueces. La prueba de la hipótesis de precognición obtuvo resultados mejores que el azar (32 % de tasa de éxito ) en la medida de resultados directos. Los artefactos metodológicos obvios se descartan y la discusión concluye con una exploración de si las evaluaciones de los jueces también apoyan la hipótesis del sueño precognitivo.

French

REVERIE PRECOGNITIVE : RECHERCHE SUR LA COGNITION ANOMALE ET LES FACTEURS PSYCHOLOGIQUES

RESUME : L’étude en ligne de précognition en rêve a examiné des variables psychologiques et parapsychologiques reconnues pour leurs contributions aux expériences de précognition en rêve. Cinquante participants ont contribué chacun à quatre essais, où la tâche consistait à rêver d’un vidéo-clip qu’ils visionneraient plus tard. Des juges indépendants furent employés pour évaluer la correspondance entre les rêves et le lot de cibles. Aucun soutien ne fut découvert à l’hypothèse selon laquelle les individus qui sont tolérants à l’ambiguïté relèteraient davantage de correspondances entre leurs rêves et les clips visionnés ultérieurement. Une relation fut découverte entre la confiance à priori des participants quant à la similitude entre leur rêve et la cible futur et la véritable similitude entre la cible et les rêves ; toutefois, il n’y avait pas de relations entre les similitudes perçues et les évaluations par les juges (taux de succès et similitudes). Le test de l’hypothèse de la précognition obtient un score supérieur au hasard (taux de succès de 32 %) sur la mesure planifiée du succès direct. Certains artefacts méthodologiques évidents sont écarts, et la discussion conclut en explorant la capacité des évaluations par les juges à soutenir également l’hypothèse d’une précognition en rêve.
OBITUARY

ROBERT L. VAN DE CASTLE
1927–2014

BY ERLENDUR HARALDSSON

Dr. Robert L. Van de Castle passed away on January 29, 2014, at the ripe age of 86. To me he was a very memorable man. I met him first at the Spring Review Meeting at the Foundation for Research on the Nature of Man (Institute for Parapsychology) in Durham, North Carolina during the year that I spent there in 1969–1970. Rhine held review meetings twice a year and invited a dinner speaker from outside. This time the speaker was Robert (Bob) Van de Castle, who had in the mid-50s spent some time with Rhine at the Duke Parapsychology Laboratory. There he investigated personality correlates of PK performance and was a research assistant to Gaither Pratt in his project for the military involving homing behavior of pigeons.

After Bob’s talk at the review meeting, we got into a conversation. This conversation had a decisive effect on my life, for which I will always be grateful. He invited me to join a 1-year internship program in clinical psychology that was being established at the Department of Psychiatry at the University of Virginia in Charlottesville. He also guaranteed me time to work on my doctoral dissertation that I was writing with Professor Hans Bender in Freiburg, Germany. That was the beginning of a long friendship. Whenever I visited Charlottesville in the ensuing years I always made it a point of seeing Bob and to spend some time with him.

Bob was born in Rochester, New York, on November 16, 1927, as the son of Omar Van de Castle (translates to “Omar from the castle”), who was born in Belgium, and a Canadian mother. He studied at the Universities of Syracuse and Missouri and obtained his PhD in Clinical Psychology from the University of North Carolina in 1959. A turning point in Bob’s life was his pioneering work with Calvin Hall at his Institute of Dream Studies in Miami, Florida. Their joint book *The Content Analysis of Dreams* (Hall & Van de Castle, 1966) became a classic on quantitative research on dreams. It established norms and revealed prominent differences in the content of dreams between men and women, just to mention one of their findings.

When Bob was invited to join the Department of Psychiatry at the University of Virginia, he was offered the opportunity for further research on dreams and sleep, and the Sleep and Dream Research Laboratory of the University of Virginia Medical Center was established with Bob as director. There I took part in weekly research meetings and found it an enriching experience to follow what research Bob and Peter Hauri—his associate—were doing at the time.

Apart from his clinical work, Bob’s life was divided between his research and writing on dreams and on the paranormal, of which the dream part was much greater. He was the author of *Our Dreaming Mind* (Van de Castle, 1994), which was described as a “landmark” by Monte Ullman, a “masterpiece” by Henry Reed, and “a sweeping compilation unsurpassed in the literature for its scope” by Stanley Krippner.

Bob published over a hundred papers in peer-reviewed journals, wrote articles in many leading newspapers and popular magazines, and discussed dreams on national TV and radio shows. He felt it important to familiarize the public with research on the fascinating realm of dreams. He played a prominent role in the International Association for the Study of Dreams, was their president in 1985, and presented regularly at their conferences, both in the US and abroad. He conducted numerous workshops on dreams on both sides of the Atlantic.

I read in his bibliography that his first publishing venture—on May 14, 1950—was in no less a newspaper than *The New York Times*. The title of the article was “Honeymoon Abroad: Cycling and Hitch-hiking Helped to Stretch $100 All the Way Across Europe” (Van de Castle, 1950). Bob was always an enterprising man; he travelled widely and knew how to enjoy life.

For the readers of the *JP*, Bob’s involvement with the paranormal will be of particular interest. We know of the time he spent with Rhine at the Duke Parapsychology Laboratory. Indeed his next four papers after “Honeymoon Abroad” stem from 1953–1959 and were published in the *JP* and the *Journal of the American Society for Psychical Research*. They all dealt with some aspect of the paranormal, from correlates to ESP performance to the sheep-goat scale.
Bob became well known for his study of psi abilities among the Cuna Indians on the San Blas Islands off Panama. In the 1960s he made several trips to the islands, where he tested a large number of Cuna adolescents in a school setting with traditional ESP cards that he had altered to make them more interesting and relevant to the Cunas: namely, a jaguar in a jungle setting, an underwater view of a shark, a conch shell on sand, a large canoe with a sail, and a propeller airplane in the sky. All the cards were in natural colors. Over 900 runs of 25 cards were administered in 1968, 1969, and 1970. The girls obtained a mean score above chance in each of the 3 years, resulting in a CR (z) of 3.67, which is highly significant. The boys scored slightly below mean chance expectation every year. The difference between the sexes was significant at the .0002 level. A possible explanation for the difference in scoring was that the girls appeared to be more cooperative and attentive than the boys. I still recall vividly when Bob reported on his work with the Cuna Indians in his Presidential Address at the 1970 Parapsychological Association convention at the Barbizon-Plaza Hotel in New York.

A notable contribution was his chapter “Sleep and Dreams” in Wolman’s *Handbook of Parapsychology* (Van de Castle, 1977). Bob’s last paper on the paranormal was “Dreams and ESP,” which he presented at the international conference of the Parapsychology Foundation in Utrecht in 2008. Its theme was Charting the Future of Parapsychology (Van de Castle, 2009). The convention was held to commemorate the first post-WWII conference on parapsychology in Utrecht in 1958.

Bob was not only a researcher and tester of subjects. Twice he acted as a receiver in dream telepathy experiments and obtained positive results on both occasions: first with Calvin Hall (1967) and later with Montague Ullman and Stanley Krippner at the Dream Laboratory of the Maimonides Medical Center in New York (Ullman & Krippner, 1970).

In the Maimonides experiments, the participants slept in the laboratory, and their rapid eye movements (REMs; indicators of dreaming) and EEG data were recorded. An agent elsewhere in the building was informed when the participant started a REM period. On each of eight experimental nights with Bob, the agent opened by a random procedure one of eight envelopes containing the target, looked at it, thought about it, and tried to interject this target into Bob’s dreams. After 10 min of each REM period, Bob was awakened over an intercom by the monitoring experimenter. He was asked what he had been dreaming, to describe it in as much detail as possible, and then go back to sleep. Upon awakening in the morning he was given the eight targets and asked to rank how close each of them was to his dreams. For the eight nights the results were highly significant, eight hits and no misses (p = .004). An additional ranking made by three judges was also highly significant (p = .001). Bob became the star subject at Maimonides.

Bob was an active member of the Parapsychological Association, a council member for 10 years, and its president in 1970.

Robert Van de Castle led a remarkable and productive life. He was easily approachable, treated everyone as his equal, was a very likeable man, and indeed a great friend.

References


*University of Iceland*

*101 Reykjavik, Iceland*

*erlendur@hi.is*
BOOK REVIEWS


Harvey Irwin has excellent credentials for writing a guide to education in parapsychology, having taught such a programme at the University of New England in Armidale, New South Wales, since the late 1970s (he retired in 2003) and having authored (lately co-authored) the highly influential textbook An Introduction to Parapsychology, now in its fifth edition (Irwin & Watt, 2007). Notwithstanding this publication, and one or two others aimed at providing an overview for the student (e.g., French & Stone, 2013; Holt, Simmonds-Moore, Luke, & French, 2012), Irwin is correct in identifying a need to provide accurate and practical guidance for education in parapsychology. This is not only to provide a pathway for the training of the next generation of parapsychologists but also to better educate their “mainstream” peers. Too much time is devoted by parapsychologists to arguing with intractably established sceptics rather than focusing on those whose views have not yet ossified. As Max Planck (1949) famously observed, “A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it”(pp. 33–34). This slim volume is divided into two sections, with Part 1 oriented to the student’s concerns and Part 2 to the tutor’s. Part 1 begins with a chapter on “Misconceptions and Preconceptions” that recognises (based in part, it seems, on the author’s personal experience) that prospective students can be drawn to courses in parapsychology with “wildly inaccurate” ideas about what the subject entails, such that they can easily become dismayed on discovering its actual scientifico-mathematical nature. Here Irwin offers a rather narrow definition of parapsychology that privileges the experimental method and the laboratory setting, which may be slightly out of kilter with a discipline that is witnessing a revival of interest in field work and qualitative methods of discovery and analysis. He notes that some phenomena may be included under the banner of parapsychology for political purposes, as a means to undermine credibility by association with dubious phenomena such as crop circles and Bermuda Triangle disappearances. Irwin concedes that their inclusion might be justified on the grounds that they reflect “mysteries” that resist conventional explanation, but this seems unnecessary given parapsychology’s origins in psychical research that clearly set out the scope, as expressed in the inside front cover of the Journal of the Society for Psychical Research, “[to investigate] without prejudice or prepossession and in a scientific spirit those faculties of man [emphasis added], real or supposed, which appear to be inexplicable on any generally recognised hypothesis.” Whatever the ultimate explanations of crop circles and the Bermuda Triangle mystery, it seems highly unlikely to expand our understanding of human capabilities. In my view parapsychology should properly distance itself from the kinds of topic that feature in Tobacyk’s Paranormal Belief Scale (including astrology and UFOlogy) and thereby deny the relevance of much of the research on correlates of paranormal belief, which has been based on such flimsy measures of superstitiousness. At the same time, Irwin argues against the inclusion of mystical and occult practices such as Tarot, I Ching, and magic in the sense of spell-casting. To be sure no respectable education programme in parapsychology would include any kind of apprenticeship or training in these methods, but insofar as success with these can be interpreted as involving psi they would seem to me to be fair game (indeed, I have collaborated on a number of projects that have used Tarot and I Ching protocols as means of testing for PK, and am currently looking at Pagan spell-casting as a possible form of noncontact healing—Martin, Drennan, and Roe, 2010; Roe, 1994).

Irwin notes that registration onto a parapsychology course might reflect personal tendencies toward magical thinking or religious convictions that variously promote or prohibit psi phenomena. The latter can include belief in forms of materialism that deny psi phenomena in principle. He labels this position skepticism but it could more accurately be termed counteradvocacy, as the former implies doubt rather than denial. Counteradvocates could possibly sign up for a parapsychology course with the aim to debunk the findings or question the adequacy of the
methods, though I suspect a rational debate on either of these would be welcomed by most tutors, who would be relatively confident of their soundness.

Chapter 2 is concerned with students’ motivations for studying parapsychology so as to ensure a good fit with content and orientation. The first of these motivations is to understand their own experiences, particularly if they have concerns about their implications; for example, are they morally bound to act on apparent premonitions? Although it is extremely difficult to account for individual experiences after the fact, courses can provide a context that helps normalise them. This is more challenging where the focus is wholly on laboratory studies, and some consideration of more phenomenological approaches—with appropriate critical reflection—would have been welcome. The second motivation for students is to find someone who understands them and their experiences, seeing the tutor as a kind of therapist. Irwin raises the fear that providing evidence (in other contexts) of psychic functioning may then reinforce pathological beliefs and behaviour. Thirdly, students may take the course in the hope that it complements other, more vocational, subjects, such as providing a grounding in the scientific method for nonspecialists. Irwin argues that a premium is to be had if parapsychology is taken as part of a psychology programme that would have already introduced many of the phenomena of social, cognitive and abnormal psychology that overlap with paranormal experience. One could similarly argue for its inclusion as part of a physics programme. The fourth motivation is to become a professional parapsychologist, though this ambition is stymied by the extremely limited job opportunities, and Irwin wisely proposes that at the least graduate studies should include some overlap or have some bearing on more mainstream subjects so as to provide an opportunity to demonstrate one’s academic credentials.

Chapter 3 begins by addressing the pros and cons from the student’s perspective of the different methods of studying parapsychology. The distinctions may seem quite self-evident to experienced academics but are usefully pointed out for the prospective student. Self-directed reading offers a convenient approach that will reflect the individual’s main areas of interest so should be motivating, but because it is unguided there is no guarantee that what is being read is of good quality or is representative of the field. Guided reading lists address this, though such lists tend to be introductory so present less of a challenge to the enthusiast. Community or Adult Education courses have a similar disadvantage and can vary widely in quality (depending on the tutor’s training or background) but at least are accessible, generally affordable, and provide the opportunity for support and stimulation from like-minded peers. Study at the bachelor’s level would provide more of a challenge and would typically be included in a broader syllabus that could even act to camouflage one’s true interests if one is concerned about the effects upon potential employers while still enabling the subject to be studied critically and in some depth. Irwin’s advice to students looking to choose a master’s or PhD programme also emphasises the need to cover one’s tracks. At one level this seems like sensible (if timid) advice that acknowledges the antipathetic climate experienced by some. However, it may actually contribute to that climate by reinforcing the stereotype that there is something unwholesome about parapsychology, and this could act to deter bright, interested students from throwing their hat into the ring. Generally, Irwin’s concerns about possible stigmatisation seem overplayed, especially where it can be made clear that parapsychology is the study of anomalous experience and so encompasses naturalistic explanations where they are appropriate.

The next section considers how to study parapsychology without a scientific background. The emphasis here is on mathematical competence, but this could be a deterrent to the many who have developed math anxiety; it might be more fruitful for the lay person to concentrate on case studies as a vehicle for exploring the pitfalls of human perception and inference and so develop a more scientific orientation to evaluating evidence that doesn’t hinge on mathematical sophistication. This section notes that the principles of the scientific method can be straightforward and intuitive, contrary to the stereotypical view of science as necessarily off-puttingly complex or requiring Einsteinian reasoning powers to be mastered. Educational programmes allow these principles to be more easily grasped through practical demonstrations (as participant, researcher, or both) and Irwin usefully offers advice on how one can gain hands-on experience either as a student or a volunteer.

In Chapter 4 he considers the various benefits of formal study, including that it can cultivate an appreciation of the (socio-political) complexities of conducting research on controversial topics. It can also provide a more solid knowledge base from which to be a consumer of the exaggerated claims (both pro and con) that are typical of the media and popular literature. The laudable assertion (p. 47) that paranormal claims should be evaluated on the adequacy of the empirical evidence seems to contradict his initial point that evidence is only part of a broader socio-political schema (which seems a more accurate description of how science is actually practised). The third benefit he identifies is an appreciation of how science works, and this seems to be an attempt to reconcile benefits
one and two. He draws attention to the fuzzy nature of science and the general dominance of theory over observation in the natural sciences that parapsychology seeks to emulate. The latter obviously suffers in such comparisons, but this may be because the comparison is inappropriate. Parapsychology seems to share more with other social sciences, which also have an ambivalent relationship with theory, reflecting in part the complexity of a subject matter that includes sentient organisms with their own intentions and expectations that create a myriad of additional extraneous factors. Parapsychology seems to be especially susceptible to such expectancy effects so that demands, for example, for replication on demand, seem excruciatingly naive. As an antidote to this potential doom and gloom, Irwin ends this chapter with a final section on “discovering the joys of being a parapsychologist” that promotes the opportunity to explore such a rich and diverse subject that, at least potentially, has something fundamental to say about the human condition.

Part 2 of the book focuses on instructor perspectives and is slightly shorter than Part 1, consisting of just three chapters. The first of these is likely to prove of most benefit (particularly to the nonspecialist academic looking to develop an interest in parapsychology) and explores the issues in designing a parapsychology course. Alternative ways of delineating the topic are explored, with an emphasis on research that utilises the experimental method and is organised around the triumvirate of ESP, PK, and survival (with perhaps an additional consideration of factors affecting paranormal belief). There is a nice symmetry here with earlier chapters, in that the discussion of how topics are demarcated shows an awareness of the need to be sympathetic to student expectations. Nevertheless, in the next section Irwin considers the need to maintain academic standards in the face of student demands for the popular or sensational. This is not a trivial issue in a time when even university-based courses use student assessments as part of the metric by which programmes are evaluated and university national league tables are produced. Mitigating against this pandering to vox populi is the need to meet learning outcomes often set at a national level that expect a degree of scholarship and critical thinking as indicators of academic quality.

In the next section Irwin offers a philosophy of teaching and learning, drawing a distinction between the more traditional teacher-centred approaches and more modern student-centred approaches. The former are typically didactic, as in a formal lecture, and are mainly focused on conveying prespecified information identified by the lecturer. The latter approach encourages individual exploration and self-discovery in which the teacher becomes facilitator. Politically, the latter has tended to supersede the former, but in practice the two approaches tend to be used complementarily, with more didactic teaching providing the platform (and setting boundaries) for personalised exploration. Of course the structure and method of delivery of a programme depends on its aims and objectives. In the next section Irwin considers these and makes some suggestions for how they might be framed. Next he looks at the context for the course, particularly in identifying opportunities to capitalise on overlap with other subject areas (such as the psychology of perception and memory), not only to build on that prior learning but also to affirm the relevance of parapsychology for what might be the student’s primary or vocational interest. Even where the student might have career aspirations in parapsychology, as already discussed, it is important to show how this work is relevant to more mainstream concerns.

Chapter 6 reviews the practicalities of teaching parapsychology. First Irwin tackles the delicate issue of an academic who has been teaching more mainstream topics for some time and who now applies to include a parapsychology element to the curriculum. Some objections to such an application can be overcome by acknowledging the subject’s controversial nature and the need to study it scientifically, and by noting its popular interest so that it provides an ideal vehicle for developing critical thinking skills among a large and enthusiastic cohort. In discussing teaching materials he encourages the sharing of good practice and of resources (including audiovisual aids). In the UK this has been an aim of a special interest group for anomalous psychology and parapsychology within the Higher Education Academy, but this could, as Irwin suggests, be done under the auspices of the Parapsychological Association.

He devotes another section to student misconceptions, but although this focuses on practical aspects it does seem repetitive of earlier coverage. Among these is the possibility that students become dismayed by the emphasis on methodology and statistics which might appear “excessively pedantic” (p. 81) but, as he explains, is essential if the conclusions we reach and the recommendations we give to the general public are to be trusted. Similarly there may be a mismatch between the findings in the research literature and the student’s own experience. This can be a source of tension, but I have found that it can be managed productively if students are encouraged to enter into a reciprocal relationship where “academic research” provides the lens through which they critically evaluate their own beliefs and experiences, and their own beliefs and experiences are used as the lens through which they evaluate the
validity of “academic research”. Similarly, Irwin goes on to argue for the importance of taking a practical approach so long as the demands placed on students are in step with their skill development—it wouldn’t be sensible to have them working as independent researchers on projects until they have gained a lot of experience of the principles and pitfalls of research. Again, many of the suggestions in this section seem sensible but unsurprising, though of course they might be much less obvious to someone just starting out as an educator in parapsychology.

In a final chapter, attention is turned to management issues, and Irwin reasonably argues that there should be some mechanism in place to establish and maintain standards, and the Parapsychological Association is brought to task for not providing this. He asserts, “the Parapsychological Association, as the professional body representing the interests of scientific parapsychology, should take on the responsibility of approving parapsychology courses that demonstrate due academic rigor” (p. 93).

The PA website does provide links to university-based educational opportunities in parapsychology around the world but stops short of endorsing these or conducting any kind of appraisal of standards. There may be difficulties in making such recommendations, particularly if any omissions are legally challenged, but these should not be insurmountable if there are clear criteria that have to be met before a course can be recommended (e.g., that a proportion of tutors are full members of the PA, that the university is nationally accredited, etc.). Irwin adds requirements for tutors to have a PhD and have published in parapsychology to demonstrate methodological competence, but both are included in the criteria for full membership of the PA so are already covered in the above. He goes on to suggest a certification system for instructors administered by the PA, but it seems more practicable to make the course the unit of assessment. Given the lack of consensus as to the core phenomena of parapsychology (textbook coverage apart), Irwin rightly eschews any assessment based on the curriculum.

In summary, this is a useful addition to the educator’s library, filled with common sense and helpful tips that would be of particular use to the early career parapsychologist who is looking to develop a teaching profile. The book mainly reflects Irwin’s experience of teaching for many years, so it does not have an exhaustive reference list but does include 74 publications that give a sound introduction to education in parapsychology.

References


Chris A. Roe

Psychology Division
University of Northampton
Park Campus
Northampton NN2 7AL, UK
chris.roe@northampton.ac.uk


Conversations With Ghosts by Alex Tanous and Callum Cooper is a short compilation of chapters, photos, interviews, and stories all centered around the somewhat legendary figure of Alex Tanous, best known to parapsychologists for his work with Karl Osis at the ASPR in New York during the 1970s and 1980s. It appropriately begins with a brief introduction by Cooper explaining who Tanous was and how the book came about. Those wishing to
know more about Tanous’ private life may feel a bit disappointed at the brevity of personal information presented here—readers are referred to Tanous’ autobiography for those details. Instead, Cooper focuses on Tanous’ qualifications and the types of research in which he participated. This sets the stage for the six chapters that represent the body of the book, which are either transcriptions of Tanous’ notes and interviews or were written by him for an unfinished set of case reports.

The first chapter introduces the reader to the idea that not all hauntings are alike. In addition, Tanous shares his thoughts on why hauntings may occur in some cases and not others and describes the manner in which the ASPR would go about choosing and investigating cases. For some of us, this is something of a walk down memory lane. Tanous, as an extremely gifted psychic, played a key role on the team. Data collection was only the first step in a case. After they gathered that information, the team would figure out what was needed to resolve the situation to the mutual satisfaction of both the living and the dead. This varied greatly from case to case but often involved letting spirits tell their stories and providing counseling to all parties as needed.

The next three chapters are arguably the most interesting. They describe Tanous’ first-hand experiences at Cedar Rapids, Hawk Mountain, Dandy House, and other haunted locations. This provides intriguing insight into how Tanous worked and the ways in which he made sense of his experiences. For example, the fact that Tanous took on the viewpoints of the spirits who told him their stories could explain why he stated that “There are no truly evil ghosts; any unpleasant effects of the presence are only the results of an unhappy lingering energy resulting from the act of injustice or imbalance itself” (p. 8). After all, every person is the hero of his or her own story! Tanous also discussed what he called a “spiraling effect” where “multiple manifestations converge on a site, drawing to the house people with similar character weaknesses as the original inhabitants who had begun the cycle” (p. 11). We might think of this as a kind of resonance, where the living are (perhaps unconsciously) drawn to a certain energy while others are repelled by it.

The fifth chapter represents a jarring stylistic shift compared to the earlier, more personable ones. This is because it is the edited transcript of an interview Tanous gave on his work in 1981. In it, he touches on everything from ghost ethnicity and types of manifestation to his utter lack of belief in demons and possession to survival research and what part of personality, if any, survives death of the body. Tanous believed every person was comprised of three things: a body, a spirit, and a soul. He refers to the soul as “individuality,” whereas the spirit part of us that can leave the body without us dying (i.e., our conscious awareness) he called “personality” or “the apparition” (pp. 72–73). One gets the feeling he is struggling to give us a glimpse of the ineffable but falls short due to the inadequacy of language to capture its complexity and nuances. The reader might be forgiven for wishing for a more comprehensive explanation of what all this meant to Tanous or at the very least a good glossary.

The last chapter returns to the style of the first ones, with a few words about Tanous’ and others’ beliefs on what part of personality, if any, survives the death of the body. It also gives the book a chance to explain how his philosophy on this came about through his experiences going out-of-body.

The book offers a diverse set of appendices towards the end. These include a previously published interview Loyd Auerbach did with Tanous, a few pages on the Amityville case—which the ASPR quickly recognized as a hoax—including a letter written by Tanous to correct the record on things falsely attributed to him, Jennifer Allen’s memories of some experiences she had with Tanous, and an edited article by Thomas Verde on Tanous’ investigative exploits on haunting. The end matter consists of brief bios of the authors and an index.

It may be no surprise, given the variety of wide-ranging material cobbled together, that the result is something of a mixed bag. Some topics, such as the overview of Buddhist, Jewish, and Christian beliefs on survival, are oversimplified to the point that they border on being insulting. Other concepts are confusingly presented, such as when the book first discounts adolescent-fueled poltergeist activity as “not representative of true manifestations” (p. 2) and then goes on to note, “Overwrought emotional states can contribute to the manifestation of phenomena with similar effects to those caused by adolescents, with regards to PK and/or poltergeist activity” (p. 2). This sounds as if the book is saying that mind-matter interaction caused by nonadolescents is only comparable to, not the same as, those caused by adolescent poltergeist agents. A notation here could have clarified what was meant and/or made sure readers knew that poltergeist agents can be of any age, that agents need not be individuals but also can include an entire group or family, and that mind-matter interaction can at times be caused by discarnate spirits, whether alone or in combination with the living.

More than once, the book’s terminology gets in the way of clarity, whether because of implicit assumptions that we no longer accept as necessarily true or because the meanings shift from one part of the book to another.
For example, the word “apparitions” is used to refer to discarnate entities who can interact with the living. Why does this matter? Because we now know that apparitions can be poltergeist effects and not involve spirits at all. However, at least the word means the same thing throughout the book. The same cannot be said for the term “ghost.”

The first chapter notes that ghosts can be in more than one place at the same time and compares this to how the living can be seen in two places at the same time during out-of-body experiences. The idea that ghosts have awareness is reinforced by the comment that “ghosts exist because they must tell their story to someone who can set their lives and actions into balance” (p. 5). Yet, later we are told that ghosts are “a replay of the incident that happened, or someone seeing someone moving, yet there’s really no dialogue” (p. 74). This sounds less like a conscious spirit than what modern parapsychologists would call place memory. Such memories are imprinted on objects and the environment by the living. The dead have nothing to do with causing place memories, and one needn’t be dead for them to be seen by others. So, if ghosts can be in two places at the same time, does this mean a conscious spirit is seen in multiple places at the same time? Or is the place memory simultaneously experienced at multiple locations? Which is it? It could make a big difference.

In conclusion, regardless of whether one agrees with Tanous’ sometimes dogmatic beliefs, it is clear that he was ahead of his time in using the same psychological counseling for spirits as he did the living. Some of the case reports are truly gripping and offer an intriguing look at the man’s inner thoughts and techniques. Yet at the same time, the material is also somewhat dated. Our understanding of hauntings and mind-matter interaction has become richer and more complex in the decades since he died. This book should not be taken as an authority on hauntings and the afterlife so much as an interesting way station from the past that allowed us to advance to where we are today. Read in that light, it becomes a useful and interesting contribution to the literature.

PAMELA HEATH

P.O. Box 6409
Alameda, CA 94501, USA
pam@pamelaheath.com


Readers of the Journal of Parapsychology will not find very many of the studies and experiments mentioned in Randi’s Prize: What Sceptics Say About the Paranormal, Why They Are Wrong and Why It Matters unfamiliar, and he does not even attempt to shed any sort of new light on them. As the title suggests, the focus is much more on the skeptics, what they do and why they behave as they do. Examples are taken from séance days to the modern era, from Harry Houdini to James Randi. The title is a little misleading, as one would expect that it would focus on Randi’s famous “million dollar challenge” and it really does not. The focus instead is on the actions and psychology of the skeptics.

McLuhan presents himself at the outset as someone who knows very little about scientific parapsychology, someone who more or less accepts the “debunkings” of Randi and others. The viewpoint of the skeptic, which he notes that in general is the viewpoint of mainstream science, seems logical enough. But, as he explains in the introduction, he cannot help but note that apparent instances of psi are not at all rare, and he also finds himself wondering why the debate is, as he puts it, so “shrill,” why Richard Dawkins refers to the paranormal as “bunk” and its advocates as “fakes and charlatans,” why Randi uses terms like “woo-woo” and considers the study of it “farce and delusion.” With this as background, McLuhan is startled when he begins to read the scientific literature of parapsychology; he is startled by the volume of it, by the numbers of incidents and experiments described, and by the level-headed and objective writings of the investigators and researchers themselves, who appear as serious as any other scientists in any other field—in other words, not even close to the descriptions of them presented by writers such as Dawkins and Randi.

Beginning in chapter 1, “Naughty Adolescent Syndrome,” he discusses several well-known poltergeist cases—as after expressing surprise that there is any current interest in such things. In particular he devotes considerable space to the Tina Resch case, which involved James Randi—and in which the defining event, at least as far as public
perception was concerned, was a photograph of a telephone flying across and in front of a startled-looking Tina. Denied access to the house and to Tina, Randi persuaded the photographer to allow him to see a number of unpublished photos from the same series, and then he announced that Tina was in such a position that by yanking on the cord she conceivably could have caused the telephone to fly in the manner seen. Additionally, he notes, on at least one occasion Tina was observed to knock over a lamp and then feign surprise, an act she admitted to when confronted. McLuhan notes that this seems like an effective “debunking” to him, but when he pursues the matter further he finds that many more events were documented in this case by William Roll, a large number of which could not be explained by tricks as simple as yanking a phone cord or upsetting a lamp. These Randi ignores, preferring to “explain” only a few and suggesting that Roll was not a careful or competent observer. This sets something of a pattern for a large number of other skeptic vs. researcher accounts that McLuhan presents in detail as the book progresses.

Also in this chapter, McLuhan notes that many cases have been “solved” when the perpetrator confesses, but also notes that—with the notable exception of the Fox sisters, which he later discusses in some detail—the “confession” is often second or third hand and is not acknowledged by the supposed perpetrator, who at times vigorously denies there ever was such a confession. A notable example is to be found in the case of the Miami poltergeist. A police sergeant told a reporter that Julio Vasquez, the young man at the center of the disturbances, had confessed to causing them by trickery. On reading this in the newspaper, Vasquez confronted the officer and told him it was a lie, that he’d said no such thing, and the officer did not deny it. McLuhan suggests that accepting these second-hand confessions at face value is convenient for the skeptics, allowing them to dismiss the case without further comment. But McLuhan finds this approach inadequate. For example, in a much older poltergeist case from Stockwell, England, the young girl at the center of the matter was said to have confessed to causing the mischief by such actions as attaching horse hairs to dishes and such to make them appear to fly. Exactly how a servant girl barely out of her teens mastered the arts of stage magic well enough to fool people who were presumably watching her—she was, as is commonly the case, under suspicion as the cause of the disturbances almost from the outset—is not explained or even addressed by skeptic writers. The supposed “confession,” again to a third party, is for them sufficient to close the case. Consistently, as he examines cases of various sorts, McLuhan notes that although the parapsychologists often cannot prove their case, the tactic of the skeptic is simply to try to create doubt, often by distorting the facts or by focusing on small details. As the media and mainstream science are inclined to accept the skeptic viewpoint, this is often all that’s required. Almost never are viable alternatives offered, if in fact any alternative explanation at all is presented.

Early on, McLuhan presents his concept of rational gravity. Most of the skeptics, Randi especially, often suggest that people, including investigators and researchers, “jump to conclusions,” that is, presume that an event can only have a paranormal explanation—because, according to Randi, they wish it to be so. McLuhan finds the opposite to be the case, that is, people assume the explanation is ordinary, just not obvious. Only after finding that ordinary explanations will not suffice do they begin to entertain the idea of something paranormal. More than that, he suggests that people will often adjust their recollections of events to make them seem more ordinary than they actually were, again a direct refutation of the skeptics’ “faulty memory” hypothesis, which they use commonly.

The next two chapters deal with other areas of the field, such as the investigations of the physical medium Eusapia Palladino and trance mediums such as Leonora Piper. While noting that many mediums of all types were frauds and were exposed as such, he, like many others, notes that there are some, such as the two mentioned, that remain of interest. But here, again, the skeptics are shown to evade the truth at best. For example, in describing a séance involving Eusapia Palladino, Hugo Münsterberg writes that Palladino was exposed when a man caught her using her foot to perform tricks, and that at this time she gave voice to “a scream as if a dagger had stabbed [her] right through her heart [and] indicated that she knew that at last she was trapped and her glory shattered.” Later in the chapter he presents the actual transcript of that particular séance, which states merely “[Palladino] screams sharply. Reason not known.”

Chapter 4, “Uncertain Science,” deals with “strange coincidences” and also with more modern laboratory investigations of psi, from card-guessing experiments to current studies. McLuhan does not neglect the problems of scientific parapsychology, such as the commonly mentioned inability of mainstream scientists to replicate parapsychology experiments in their laboratories, and he does not fail to mention the well-known cases of fraud, such as the Creery sisters, S. G. Soal (McLuhan makes a small error here in giving Soal the initials “G. R.”) and Walter J. Levy. For some of the skeptics, even a few cases of fraud are enough to discredit the entire field. But as McLuhan points out, fraud has occurred in every area of science, and there is no reason to believe that
Book Reviews

parapsychology should have been exempt. There are some interesting sections in this chapter, in particular those concerning Rupert Sheldrake’s experiments with the sense of being stared at and the dog that knows when its master is coming home. In both cases, attempts at replications were conducted by skeptics. In the staring experiments, the skeptics studies’ initially confirmed Sheldrake’s results, whereupon they insisted they must have been an artifact. In the case of Sheldrake’s dog experiments, McLuhan spends quite a few pages describing the attempts by Richard Wiseman to replicate them (which Wiseman claimed failed), showing in the end that the Wiseman study was designed to fail and, as Sheldrake has claimed, the Wiseman data actually match his own rather closely—but do not fit the extra criteria Wiseman imposed, criteria seemingly designed specifically to cast doubt on the Sheldrake study. Also discussed at some length is the attempt by Wiseman to use statistical meta-analysis (a statistical summing of multiple studies) to discredit the work of Honorton and Bem, which had been published in the mainstream journal Psychological Bulletin. Yet again, McLuhan notes that skeptics like Wiseman are not at all hesitant to publicly announce that they have “debunked” a study, when what lies behind that may be very different from what is being presented.

In the next two chapters, McLuhan presents some of the evidence for and against various other sorts of phenomena, such as ghosts, out-of-body experiences, and reincarnation. The casual dismissal of events that seem to beg explanation by the skeptics becomes predictable and almost tiresome—and for the most part they rarely offer any viable alternative explanations. The section on the psychology of the skeptic is, however, quite interesting.

In the last chapter, he considers the changes that might take place in a world where ESP and PK are accepted as real, and he finds them to be major. He also seems to feel that the reality of psi implies the validity of survival of consciousness, and that this in turn takes us out of the realm of science and into religion. This seems excessive; the reality of psi as one more parameter of the universe does not necessarily imply survival of consciousness, and if survival were to prove to be a reality, that in and of itself does not really say anything about religion in much of any way. It should always be remembered that much of the science we understand quite well today and use on a daily basis would have seemed like magic to our ancestors.

In any event, Randi’s Prize: What Sceptics Say About the Paranormal, Why They Are Wrong and Why It Matters is a pleasant and informative read for individuals well-versed in parapsychology, and it does offer insights into the often curious behavior of the skeptics. Its importance, however, is to the lay public and perhaps to mainstream scientists in other fields. McLuhan is quite convincing in his thesis that parapsychology, although still an imperfect science, is of major importance, and the skeptics’ attempts to discredit it by any means possible are a disservice to the advancement of human knowledge. Recommended.

Graham Watkins

1011 N Driver St.
Durham, NC 27701, USA
cool@mindspring.com
EDITOR’S NOTE: POLICY ON AUTHOR PSEUDONYMS

Because of the well-known prejudice against parapsychology among academics in mainstream scientific fields, especially psychology, professors or researchers in such fields often risk reprisals from their universities or research institutes or damage to their professional reputations if they are identified as having published anything favorable to parapsychology. This is particularly damaging to our field if they are inhibited from publishing reports of psi experiments that they conducted or supervised, even in parapsychology journals. However, such individuals might be less reluctant to publish such manuscripts if they could do so anonymously. Therefore, I have made a policy decision. I hereby invite potential authors in this predicament who want to publish an empirical research report or any other class of scientific paper that we normally publish to submit their articles to the Journal of Parapsychology with the understanding that the author names will be listed as pseudonyms of their choosing if the paper is published after undergoing our standard peer-review process. There is precedent for this policy in the Journal. In 1968, Volume 32, pp. 153–166, we published a paper entitled “ESP Experiments with Mice” by Pierre Duvall and Evelyn Montredon. Although the names were not listed as pseudonyms, Pierre Duvall was actually the distinguished French biologist and entomologist Remy Chauvin, who at the time was a senior research fellow at the Sorbonne. My policy will be to identify the author names as pseudonyms in a note in the article itself but not in the Table of Contents or index. Although I insist that authors use their real names in correspondence with me, I promise not to divulge their names to anyone else. This includes the referees of the paper, who will not be informed that the names they get are pseudonyms. Let me close by saying that I hope this policy will contribute in some small way to furthering our understanding of psi by bringing material supportive of this goal to the awareness of the parapsychological and broader scientific communities.
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, unpatterned sensory stimulation; an audiovisual ganzfeld may be accomplished by placing halved ping-pong balls over each eye of the subject, with diffused light (frequently red in hue) projected onto them from an external source, together with the playing of unstructured sounds (such as “pink noise”) into the ears.

GENERAL EXTRASENSORY PERCEPTION (GESP): A 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 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.

RETOACTIVE 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. 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. 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. Table titles and figure captions should be in the text, not part of the table or graphic, and no lines should be included in tables. 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 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.