

# ‘Heads I Win, Tails You Lose’

## *How Parapsychologists Nullify Null Results*

.....

*Parapsychologists have tended to view positive results as supportive of the psi hypothesis while ensuring that null results don't count as evidence against it. Here's how this self-deceptive process works and four suggestions to overcome it.*

RICHARD WISEMAN

**A**fter more than sixty years of experimentation, researchers have failed to reach a consensus about the existence of psi (psychic ability). Some argue that there exists overwhelming evidence either for or against the psi hypothesis, while others believe that it simply isn't possible to answer the question one way or the other. One of the main obstacles to closure on the psi question involves the way in which null results are viewed (Alcock 2003). Many parapsychologists have adopted a “heads I win, tails you lose” approach to their work, viewing positive results as supportive of the psi hypothesis while ensuring that null results do not count as evidence against it.

## Cherry-Picking New Procedures

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 as evidence against the existence of psychic ability, such null findings are usually attributed to the experiment being carried out under conditions that are not psi-conducive. They are either never published (the “filedrawer effect,” see Douglas M. Stokes, “The Shrinking Filedrawer,” *SI*, May/June 2001) or are quietly forgotten even if they make it into a journal or conference proceeding. Once in a while one of these 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. In addition, 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, usually viewed by proponents as tentative evidence for psi, acting as a catalyst for further work.

To my knowledge, only one paper has revealed an insight into the potential scale of this problem. Watt (2006) summarized all of the psi-related final-year undergraduate projects that have been supervised by staff at Edinburgh University’s Koestler Parapsychology Unit between 1987 and 2007. Watt tracked down thirty-eight projects, twenty-seven of which predicted overall significant performance on a psi task with the remainder predicting significant differences between experimental conditions. The work examined a range of new and established procedures, including, for example, 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. Interestingly, Watt’s paper also demonstrated a reporting bias. Only seven of the thirty-eight studies had made it into the public domain, presented as papers at conferences held by the Parapsychological Association. All of these papers had predicted overall significant performance on the psi task. 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 percent (three out of twenty) of those that remained unreported. Watt’s analysis, although informative, underestimates the total number of psi-related studies undertaken at Edinburgh University because it did not include projects undertaken by students prior to their final year, experiments run by postgraduate students and staff, or any work conducted before 1987. Multiply these figures by the number of parapsychologists who have conducted and supervised psi

---

*Richard Wiseman is a professor in the Psychology Department at the University of Hertfordshire, U.K. Visit his Web site at [www.richardwiseman.com](http://www.richardwiseman.com) and contact him at [r.wiseman@herts.ac.uk](mailto:r.wiseman@herts.ac.uk).*

research across the world over the last sixty years or so, and the scale of the issue becomes apparent.

## Explain Away Unsuccessful Attempted Replications

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. Perhaps the most far-reaching version of this “get out of a null effect free” card involves an appeal to the “experimenter

**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-inhibited nature of the parapsychologist running the study.**

effect,” wherein any negative findings are attributed to the psi-inhibited nature of the parapsychologist running the study.

This nullifying of null findings permeates parapsychological literature. For example, Kanthamani and Broughton (1994) report a large-scale attempt to replicate the alleged ganzfeld telepathy effect, wherein one participant (referred to as a receiver) experiences a mild form of sensory deprivation and is then asked to identify a target being viewed by another person (a sender) in a distant location. Parapsychologists have employed various types of targets in these experiments, including photographs and drawings (static targets) and video clips (dynamic targets). In the studies described by Kanthamani and Broughton, the target material consisted of randomly chosen pictures (mainly postcard-sized art prints). The project involved a huge amount of work: researchers ran a series of experiments over a six-year period and conducted more than 350 individual ganzfeld sessions. The studies yielded a nonsignificant cumulative effect. However, Kanthamani and Broughton spent no time discussing whether this null finding might act as evidence against the psi hypothesis and instead simply concluded that “it is probably safe to say that static picture targets remain a less than ideal choice for ganzfeld experiments.”

Once again, this process represents the “heads I win, tails you lose” principle. Successful replications are seen as evidence

of psi, while null results are attributed to the non-psi-conducive conditions under which the replication was carried out.

### Data Mining

In addition to explaining away null findings via allegedly failed procedural modifications, some parapsychologists also adopt an “any anomaly will do” attitude and data mine in an attempt to produce some kind of psi-related result. Although such post hoc data mining might help guide future work, it has little if any evidential value. Nevertheless, parapsychologists often present it as tentative evidence in support of the psi hypothesis.

Willin’s (1996) description of his ganzfeld psi studies presents a striking example of this process at work. Willin con-

**In addition to explaining away null findings via allegedly failed procedural modifications, some parapsychologists also adopt an “any anomaly will do” attitude and data mine in an attempt to produce some kind of psi-related result.**

ducted one hundred ganzfeld sessions over a fifteen-month period, taking the unusual step of using musical clips as targets. The study obtained a nonsignificant result. However, rather than explore whether this null finding counts as evidence against the psi hypothesis, 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.”

This type of data mining again shows the “heads I win, tails you lose” principle in action, with any null effects being nullified by the apparent discovery of post hoc findings.

### Meta-Analyses and Retrospective Data Selection

After several studies have been conducted using a new procedure, parapsychologists usually carry out some form of meta-analytic review of the work. If the combined outcome of the studies is significant, the meta-analysis is usually the subject of considerable debate, with proponents believing that the finding represents evidence of psi and skeptics arguing that it may have

a normal explanation (including, for example, publication bias, inappropriate inclusion criteria, and poor methodology). However, if the cumulative effect is nonsignificant, parapsychologists often attribute this null effect to the non-psi-conducive procedural variations described in the preceding section.

Perhaps more important, the procedurally heterogeneous collection of studies usually presents parapsychologists with an opportunity to “explain away” overall null effects by retrospectively identifying a subset of studies that used a certain procedure and yielded a significant cumulative effect.

A striking illustration of this occurred in the late 1990s during a meta-analytic debate surrounding the ganzfeld psi studies. In 1999, Milton and Wiseman published a meta-analysis of all ganzfeld studies that were begun after 1987 and published by the start of 1997, and they noted that the cumulative effect was both small and nonsignificant (Milton and Wiseman 1999). Some parapsychologists criticized this analysis, arguing that they had included all of the ganzfeld studies conducted during this period and that they should have instead focused on those that had employed a “standard” procedure developed by parapsychologist Charles Honorton and his colleagues during a seminal set of ganzfeld studies conducted at the Psychophysical Research Laboratory (PRL) in the late 1980s. The difficulties with this approach became clear when researchers were unable to settle on what would constitute a “standard” set of procedures (Schmeidler and Edge 1999). Eventually, Bem, Palmer, and Broughton (2001) set out to tackle this issue experimentally, asking several people to rate the degree to which the studies in our analysis had employed Honorton’s “standard” ganzfeld procedure and then correlating their ratings against the effect size of each study. Rather than provide their own description of this “standard” procedure, Bem, Palmer, and Broughton had the raters read relevant sections in two previous papers describing the PRL studies. However, they also added a series of additional conditions, informing their raters, for example:

You should treat as standard the use of artistic or creative subject samples (as one of the most successful components of the PRL experiments used such a sample) or subjects having had previous psi experiences or having practiced a mental discipline such as meditation (as such subjects were shown to be the best scorers in the PRL experiments).

The addition of participant selection as an allegedly “standard” condition was not mentioned in the method section of either of the papers describing the PRL work. As such, it could be seen as an excellent example of retrospective data fitting, wherein parapsychologists decide which studies to analyze (or, in this instance, the weight assigned to them) on the basis of their known outcome.

Once again, it’s the “heads I win, tails you lose” principle. A significant overall effect is seen as evidence for psi while a null effect initiates post hoc searching for pockets of significance.

### Decline Effects and Jumping Ship

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 (see Kennedy 2003 for a review of this approach).

The decrease in alleged psi often causes some parapsychologists to abandon ship in search of a new procedure, placing them back at square one, ready to repeat history. This is not a new observation. For example, writing over thirty years ago, parapsychologist Joseph Gaither Pratt noted:

One could almost pick a date at random since 1882 and find in the literature that someone somewhere had recently obtained results described in terms implying that others should be able to confirm the findings. . . . One after another, however, the specific ways of working used in these initially successful psi projects have fallen out of favor and faded from the research scene—except for the latest investigations which, one may reasonably suppose, have not yet had enough time to falter and fade away as others before them have done. (Pratt 1978)

This constant “ship jumping” is one of the defining features of psi research, with new paradigms emerging every decade or so. Take, for example, the different trends in ESP research that have emerged over the years. 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.

## Conclusion

Parapsychologists have tended to adopt a “heads I win, tails you lose” approach to their work, viewing positive results as supportive of the psi hypothesis while ensuring that null results do not count as evidence against it. This involves cherry-picking new procedures from a mass of chance results, varying any allegedly “successful” procedures and then blam-

ing these variations for any lack of replication, searching for pockets of post hoc significance whenever a meta-analysis produces a null result, explaining away decline effects as an inherent property of psi, and finally jumping to the next new promising procedure. This giddy process results in an ambiguous dataset that, just like the classic optical illusion of the old hag and attractive young woman, never contains enough information to allow closure in one direction or the other.

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.

I hope that this process will help consign the psi debate to the history books and parapsychologists will no longer find themselves sitting on the fence arguing the “there is enough evidence to justify further work but not enough to conclude one way or the other” position. Rather than nullify null results, experimenters should be brave enough to give it their best shot and finally discover whether psi actually exists. □

## References

- Alcock, J.E. 2003. Give the null hypothesis a chance: Reasons to remain doubtful about the existence of psi. In *Psi Wars: Getting to Grips with the Paranormal*, ed. J. Alcock, J. Burns, and A. Freeman, 29–50. Charlottesville, VA: Imprint Academic.
- Bem, D.J., J. Palmer, and R.S. Broughton. 2001. Updating the ganzfeld database: A victim of its own success? *Journal of Parapsychology* 65: 207–218.
- Kennedy, J.E. 2003. The capricious, actively evasive, unsustainable nature of psi: A summary and hypotheses. *Journal of Parapsychology* 67: 53–74.
- Kanthamani, H., and R.S. Broughton. 1994. Institute for Parapsychology ganzfeld-ESP experiments: The manual series. Proceedings of Presented Papers: The Parapsychological Association 37th annual convention, 182–189.
- Milton, J., and R. Wiseman. 1999. Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin* 125: 387–391.
- Pratt, J.G. 1978. Prologue to a debate: Some assumptions relevant to research in parapsychology. *The Journal of the American Society for Psychical Research* 72: 127–139.
- Rao, K.R., and J.R. Palmer. 1987. The anomaly called psi: Recent research and criticism. *Behavioral and Brain Sciences* 10: 539–51.
- Schmeidler, G.R., and H. Edge. 1999. Should ganzfeld research continue to be crucial in the search for a replicable psi effect? Part II. Edited ganzfeld debate. *Journal of Parapsychology* 63: 335–388.
- Watt, C. 2006. Research assistants or budding scientists? A review of 96 undergraduate student projects at the Koestler Parapsychology Unit. *Journal of Parapsychology* 70: 335–356.
- Willin, M. J. 1996. A ganzfeld experiment using musical targets. *Journal of the Society for Psychical Research* 61: 1–17.