

“HEADS I LOSE, TAILS YOU WIN”, OR,  
HOW RICHARD WISEMAN NULLIFIES POSITIVE RESULTS:  
A RESPONSE TO WISEMAN’S (2010)  
CRITIQUE OF PARAPSYCHOLOGY

by CHRIS CARTER

INTRODUCTION

Psychologist Richard Wiseman is a well-known British critic of parapsychology, frequently appearing in the British media to ‘debunk’ psychical research. In a recent *Skeptical Inquirer* (Wiseman, 2010) article, “‘Heads I win, tails you lose’: how parapsychologists nullify null results”, Wiseman argues that parapsychologists have tended to view positive results as supporting the existence of psi, yet have adopted various strategies to ensure that null results do not count as evidence for the non-existence of psi. In this paper I shall demonstrate that throughout Wiseman’s career he has been equally culpable of adopting a “heads I win, tails you lose” approach to parapsychology’s research findings, in his case viewing null results as evidence against the psi hypothesis, while attempting to ensure that positive results do not count as evidence for it.

In his article Wiseman (2010) levels the following criticisms against parapsychologists:–

- **Cherry Picking New Procedures.** By this Wiseman means that positive findings in parapsychology have “emerged from a mass of non-significant studies. Nevertheless, they are more likely than non-significant studies to be presented at a conference or published in a journal”.
- **Explain Away Unsuccessful Attempted Replications.** Wiseman argues that parapsychologists come up with various excuses for not accepting failures to replicate positive results as evidence for the non-existence of psi.
- **Meta-Analyses and Retrospective Data Selection.** Wiseman argues that meta-analysis provides evidence against the existence of psi, but that parapsychologists retrospectively decide only to analyse data that fits with the existence of psi.

*Cherry Picking New Procedures*

In this section Wiseman (p.37) wrote:–

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 . . . they are either never published . . . or are quietly forgotten even if they make it into a journal or conference proceedings.

But how does he know that “most of these studies do not yield significant results”? He provided not a shred of evidence for these claims, yet continued:–

Once in a while one of these studies produces significant results . . . the evidential status of these positive findings is problematic to judge because they have emerged from a mass of non-significant studies. Nevertheless they are more likely than non-significant studies to be presented at a conference or published in a journal.

Again, Wiseman offered no supporting evidence for these claims. When he

remarked that “the evidential status of these positive findings is problematic to judge because they have emerged from a mass of non-significant studies”, he refers to what is known as the ‘file-drawer’ problem: that successful studies are more likely to be written up and accepted for publication, whereas unsuccessful studies are more likely to end up discarded in the researcher’s file drawer.

It has long been believed that in all fields there may be a bias in favour of reporting and publishing studies with positive outcomes. Given the controversial nature of their subject, parapsychologists were among the first to become sensitive to this problem, and in 1975 the Parapsychological Association adopted a policy opposing the withholding of non-significant data, a policy unique among the sciences. In addition, the sceptical British psychologist Susan Blackmore (1980) conducted a survey of parapsychologists to see if there was a bias in favour of reporting successful ganzfeld results, and concluded that there was none—unsuccessful studies were as likely as successful ones to be published.

Still, since it is impossible in principle to know how many unreported studies may be sitting in file drawers, meta-analysis provides a technique to calculate just how many unreported, non-significant ganzfeld studies would be needed to reduce the reported outcomes to chance levels. In a ganzfeld debate between sceptic Ray Hyman and parapsychologist Charles Honorton, Hyman had raised the possibility that the positive results were due to selective reporting. However, once Honorton calculated that the results could only be explained away by a ratio of unreported-to-reported studies of approximately fifteen to one, it is not surprising that Hyman concurred with Honorton that selective reporting could not explain the significance of the results (Hyman & Honorton, 1986, p. 352).

However, Wiseman could be described as engaging in a “cherry picking procedure” of his own, as in the Natasha Demkina Case. In September 2004 Wiseman took part in a classic debunking exercise, claiming that a young Russian girl who had seemingly psychic powers of medical diagnosis had failed a test he and his fellow sceptics designed. In fact, the girl scored at a level well above chance.

Natasha Demkina, then 17 years old, claimed that she could look deep inside people’s bodies, examine their organs, and spot when something was wrong. As part of a test broadcast on television by the *Discovery Channel*, Demkina was given a set of seven cards, with a medical condition indicated on each. Medical subjects with these seven conditions (one of which was “no condition”), each bearing an identifying number, stood in a row and Demkina had to mark each card with the number of the person who she thought had the condition indicated on the card. Under the tightly-controlled conditions imposed by the experimenters, she identified four of the seven correctly. The odds of getting 4 hits or more out of 7 by chance are more than 50 to 1 against. Another way of expressing this would be to say the probability that Natasha displayed no genuine ability but merely got lucky is less than 2 per cent.

However, Wiseman declared the test a “failure”. He was only able to do this because the experimental protocol, to which Natasha and her agent had been asked to agree, curiously states:—

If Natasha correctly matches fewer than 5 target medical conditions, then the Test Proctor will declare that results are more consistent with chance guessing and does not support any belief in her claimed abilities.

Accordingly, it was announced that Natasha had 'failed the test'. Brian Josephson, a Nobel Laureate in physics, investigated Wiseman's claims about this test and found them to be seriously misleading.<sup>1</sup> Keith Rennolls, Professor of Applied Statistics, University of Greenwich, wrote a letter that appears in the 17th December 2004 issue of the *Times Higher Education Supplement*. In part it reads:—

I have reviewed Professor Josephson's arguments, published on his web page, and find them to be scientifically and statistically correct. In contrast, the statement of Professor Wiseman, of CSICOP, "I don't see how you could argue there's anything wrong with having to get five out of seven when she agrees with the target in advance", demonstrates a complete lack of understanding of how experimental data should be interpreted statistically.

The experiment is woefully inadequate in many ways. The chance of the observed 4 successes 7 subjects by pure guessing is 1 in 78, an indication of a significantly non-random result, as claimed by Professor Josephson. . . .The experiment, as designed, had high chances of failing to detect important effects.

Here, then, we have a case in which Wiseman nullified a positive result by selectively ignoring a level of performance that is commonly accepted in social scientific experiments as significantly above what could be expected by chance alone. In other words, Wiseman 'cherry picked' an experimental design that had high chances of failing to detect important effects.

#### *Explain Away Unsuccessful Attempted Replications*

Wiseman's second criticism of parapsychological research referred to a claimed tendency for proponents to come up with various excuses for failures to replicate positive results so as to avoid accepting these as evidence for the non-existence of psi. Regarding follow-up studies of successful psi experiments, Wiseman (2010, p.37) complained:—

Any failure to replicate [the original effect] can be attributed to the procedural modifications rather than to the non-existence of psi. 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 researchers running the study.<sup>2</sup>

Again, Wiseman could be described as guilty of the practice for which he chastises parapsychology. In the highly publicized case of Jaytee, "*a dog that knew when its owner was coming home*", Wiseman attempted to explain away a potentially embarrassing *successful* replication. Jaytee's owner, Pamela Smart, claimed that the dog could anticipate her arrival home, even when she

---

<sup>1</sup> [www.tcm.phy.cam.ac.uk/~bdj10/propaganda/](http://www.tcm.phy.cam.ac.uk/~bdj10/propaganda/)

<sup>2</sup> It may strike the reader as hypocritical of Wiseman to dismiss appeals to experimenter effects, especially remarks about psi-inhibitory effects, when he was involved in one of the best documented studies demonstrating this effect. (Wiseman & Schlitz, 1997). Wiseman collaborated with Marilyn Schlitz to run identical studies in the same location using the same equipment, in order to see if participants could detect whether or not the experimenter was staring at them. Wiseman's results were not significantly different from chance, while experiments involving Schlitz produced results significantly higher than chance would predict.

returned at completely unpredictable times. It seemed as though Jaytee would begin waiting by the window at about the time she set off on her homeward journey (the following was described in Sheldrake, 1999a, 1999b, 2000).

In April 1994 Smart read an article in the *Sunday Telegraph* about research into animals that seem to know when their owners were coming home, being undertaken by biologist Rupert Sheldrake. She contacted him and volunteered to take part in his research. After receiving a grant from the Lifebridge Foundation of New York, Sheldrake began videotaped experiments with Jaytee in May 1995. Between May 1995 and July 1996, thirty videotapes were made of Jaytee's behaviour under natural conditions while Smart was out and about. Her parents were not told when she would be returning, and she usually was not sure herself. The results showed that Jaytee waited at the window far more when Smart was on her way home than when she was not, and this difference was highly statistically significant ( $p < 0.000001$ ).

The researchers discovered early that Jaytee responded even when Smart set off at randomly selected times. This was an important discovery, as it seemed to clearly rule out an explanation based upon routine, or expectations based upon the behaviour of her parents. Consequently, twelve more experiments were videotaped in which Smart returned home at random times, determined by the throw of dice after she had left her home. Figure 1 shows the results of these twelve videotaped experiments (from Sheldrake, 1999, p. 61). This clearly shows that Jaytee was at the window far more when Smart was on her way home than during the main period of her absence (55% versus 4%). The difference is highly statistically significant, with a  $p$ -value of 0.0001, implying odds against chance of over 10,000 to one.

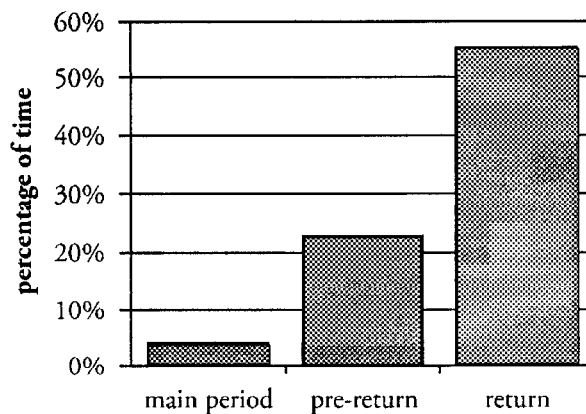


Figure 1. Percentage of time spent by Jaytee at the window when Pam returned home at randomly selected times.

The general pattern of Jaytee's response can be seen more clearly in the following three graphs (Figure 2), which summarize the average results from long, medium, and short absences. The horizontal axis shows the series of ten-minute periods (p1, p2, etc.) from the time she went out until she was on her way home. The last period shows the first 10 minutes of Smart's return journey. The graphs clearly show that Jaytee spent more time at the window

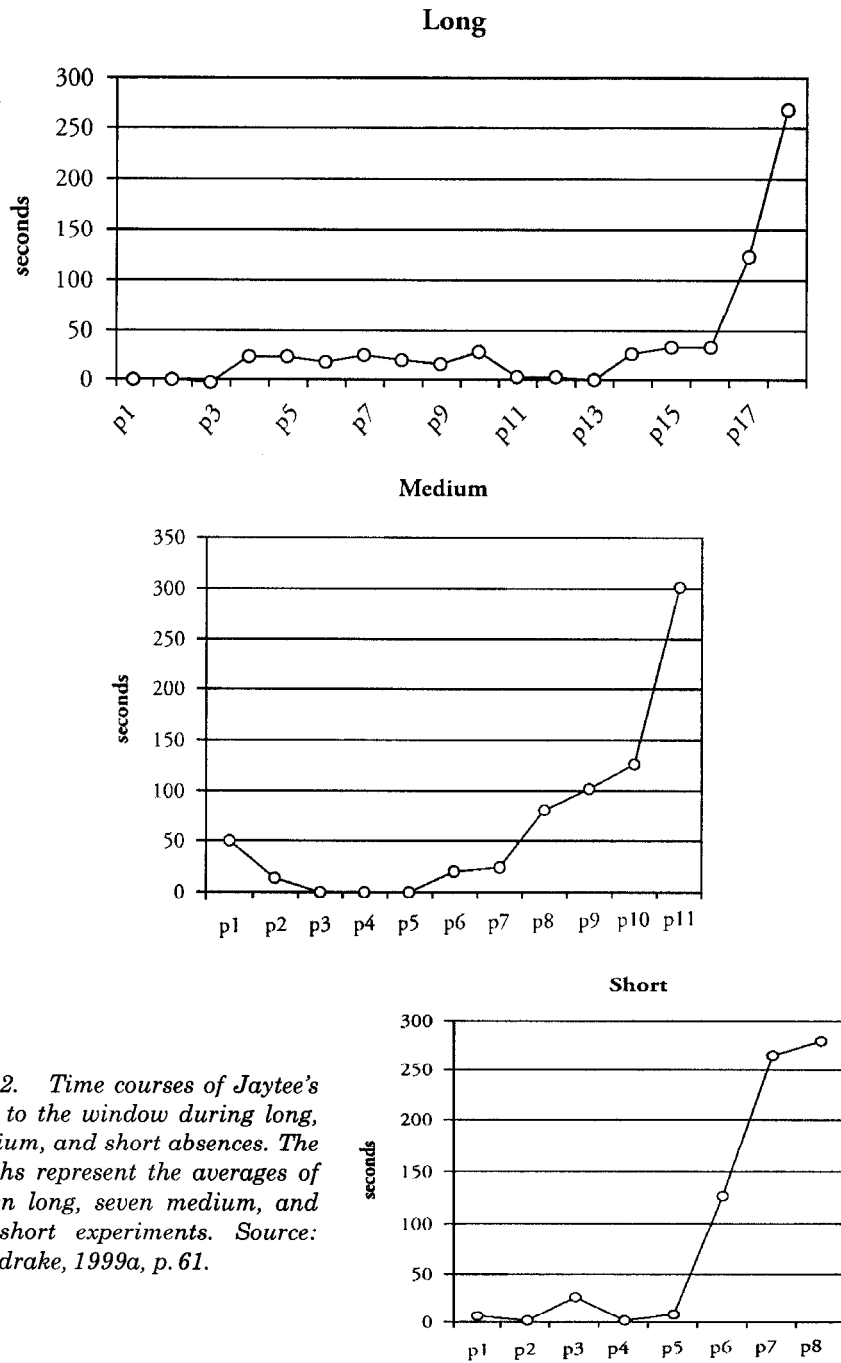


Figure 2. Time courses of Jaytee's visit to the window during long, medium, and short absences. The graphs represent the averages of eleven long, seven medium, and six short experiments. Source: Sheldrake, 1999a, p. 61.

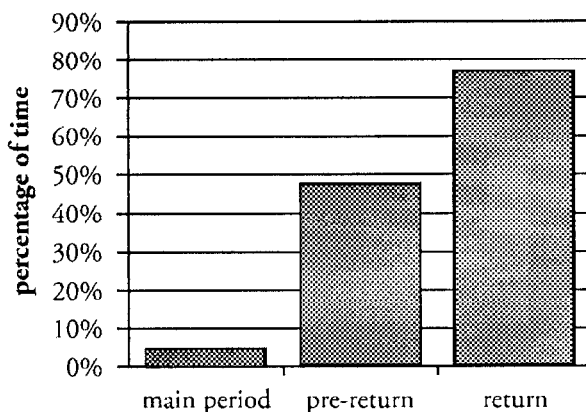
during Pam's return journey, and also that he usually started waiting at the window shortly before she set off, as she was thinking of returning.

Following the televised segment of this experiment, a number of reports about this research appeared on British and European television and in newspapers. Journalists sought out a critic to comment on the results, and

the obvious choice for many was Richard Wiseman. He suggested a number of possible explanations that Sheldrake had already tested and eliminated, such as routine times of return and selective memory. However, rather than debate the issue, Sheldrake simply invited Wiseman to perform some tests of his own. Smart and her family kindly agreed to help him.

In his four experiments, Wiseman personally videotaped Jaytee, while his assistant, Matthew Smith, went out with Smart and videotaped her. They went out to pubs or other places five to eleven miles away, and returned at times selected randomly by Smith once they were out. Smith himself knew in advance when they would be returning, but did not tell Smart until it was time to go. Wiseman, back in the apartment, did not know when they would be returning. Furthermore, Smart and Smith travelled by taxi or by Smith's car, in order to eliminate the possibility that Jaytee was listening for the sound of a familiar vehicle. Three of Wiseman's experiments with Jaytee were performed in Smart's parents' flat, similar to the experiments Sheldrake had conducted. The fourth experiment was performed in Smart's sister's flat, but Jaytee fell ill during the experiment. The results from Wiseman's three experiments in Smart's parents' flat are shown in Figure 3.

As in Sheldrake's experiments, Jaytee was at the window much more when Smart was on her way home than during the main period of her absence (78% versus 4%). With only three experiments, the sample size was small, but the results were still statistically significant, with a  $p$ -value of 0.03. In other words, Wiseman had replicated Sheldrake's results. However, much to Sheldrake's astonishment, in the summer of 1996 Wiseman went to a series of conferences announcing that he had refuted the 'psychic pet' phenomenon, and he later appeared on a series of television shows claiming to have refuted Jaytee's abilities. How did he justify his conclusions?



*Figure 3. Wiseman's results.*

Simple: Wiseman used an arbitrary criterion for success in the experiment, a criterion that enabled him to ignore most of the data he gathered. If Jaytee went to the window "for no apparent reason" at any time during the experiment, Wiseman simply ignored all the rest of the data and declared the experiment a failure. These "failures" occurred during the four per cent of the

time Jaytee was at the window when Smart was absent. After these “failures”, the rest of the data were ignored, even though Jaytee was at the window 78% of the time when Smart was on her way home.

Sheldrake met Wiseman in September 1996 and pointed out to him that his data showed the same pattern as found in the original data. Sheldrake made it clear that, far from refuting Sheldrake’s results, Wiseman’s own data *replicated* them. He even gave Wiseman copies of graphs showing him the data from his own experiments. Figure 4 shows, for instance, the graphs from the three experiments that Wiseman ran with Jaytee in Smart’s parents’ apartment.

By Wiseman’s standards, only the fourth experiment—the one performed in Smart’s sister’s apartment—was a partial success, because only in this trial did Jaytee go to the window “for no apparent reason” *for the first time* during the period when Smart was on her way home. (The videotape record showed that his visit to the window coincided *exactly* with Pam setting off on her way home.) However, Wiseman did not consider the fourth trial a success, because Jaytee did not stay there for at least two minutes, but instead left the window and vomited.

Over the next two years, Wiseman repeatedly announced through the media that he had discredited the dog’s ability to anticipate his owner’s return. For instance, on the television programme, *Strange but True*, he said of Jaytee: “In one out of four experiments he responded at the correct time—not a very impressive hit rate, and it could just be a coincidence” (ITV: 1 November 1996). The three ‘misses’ are the experiments summarized in Figure 4.

Wiseman dismissed Sheldrake’s graphical analysis of his data, calling it “post hoc”, implying that it is somehow unscientific to analyse graphically data someone else has collected. However, it is important to remember that Sheldrake applied exactly the same graphical analysis to his own data two months *before* Wiseman arrived on the scene and for two years afterwards.

As mentioned, Wiseman used an arbitrary criterion for success in the experiment, a criterion that enabled him to ignore most of the data he gathered. An analogy would be if Wiseman were to set out to test the claim that a radical new treatment is more effective in treating a form of cancer than conventional treatments; set the criterion that if any patient in the control group showed an improvement ‘for no apparent reason’ at any time during the experiment, then the experiment would be declared a failure; use this criterion to ignore the majority of his data; and then announce to the press that his experiment shows that this new treatment does *not* have a greater success rate in treating cancer, despite his own long-term evidence to the contrary.

During the controversy that followed, Susan Blackmore (1999, p.18) came to Wiseman’s aid in a newspaper article, claiming that there was a fatal flaw in Sheldrake’s experiment.

Sheldrake did 12 experiments in which he beeped Pam at random times to tell her to return. Now surely Jaytee could not be using normal powers, could he? No. But there is another simple problem. When Pam first leaves, Jaytee settles down and does not bother to go to the window. The longer she is away, the more often he goes to look.

Blackmore’s point is simply that Jaytee spends more time by the window the longer his owner is away, so that inevitably the dog will spend more time at the window in the period during which Smart returns than in any earlier

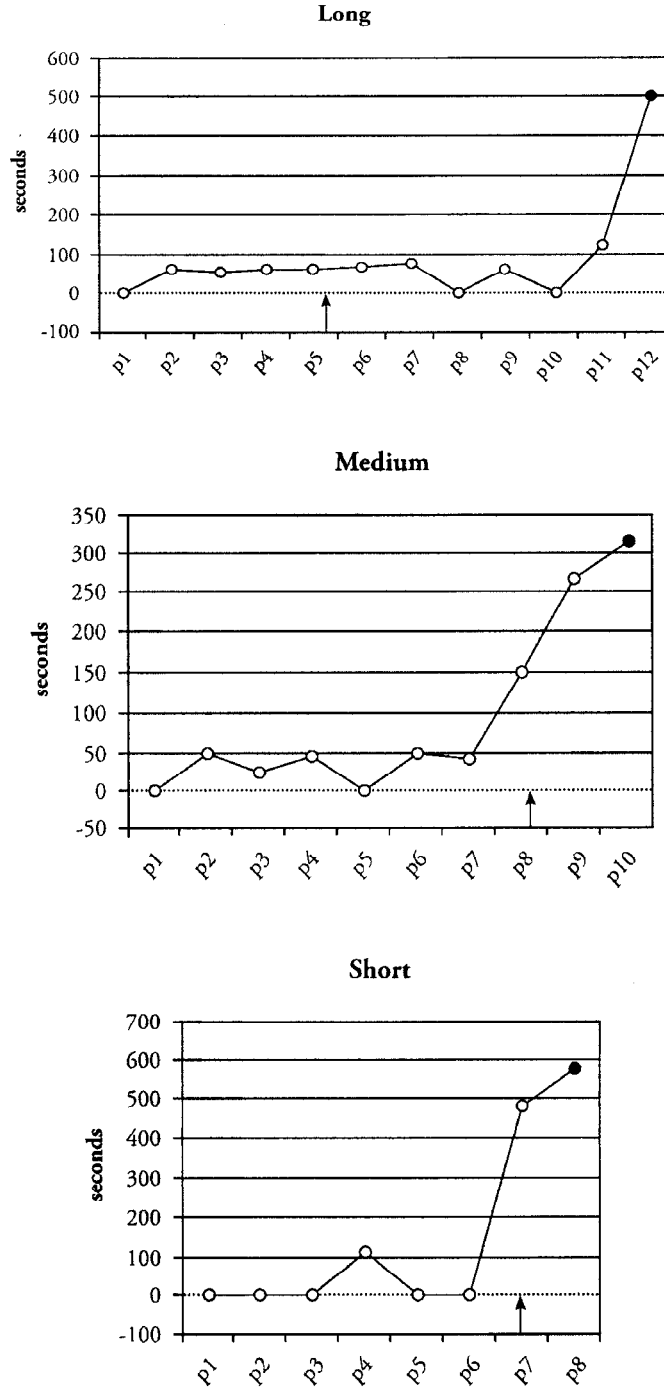


Figure 4. Wiseman's results, three experiments in Smart's flat. The periods after which Wiseman ignored the data are indicated by arrows. The final points on each graph represent the first 10 minutes of Smart's return journey, indicated by a filled circle.



period. But anyone who looks at the actual data can easily see that Blackmore's remark is simply not true. For instance, in Figure 3, we can see that, during the short absences, Jaytee spends the most time by the window when Smart is on her way home, but there is no comparable increase in time spent at the window in this same period during the medium and long absences. Likewise, the spike in time Jaytee spends by the window when Smart is on her way home during the medium absences does not show up in Period 11 of the long absences.

Sheldrake also made a series of videotapes on evenings when Smart was not coming home until very late, or staying out all night. These tapes serve as controls, and they show that Jaytee did not go to the window more and more the longer she was away. Once again, a close examination of the evidence shows the need to treat the claims of the sceptics with scepticism.

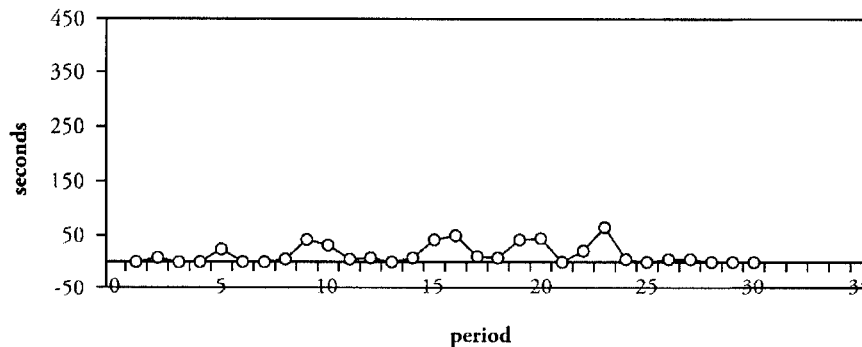


Figure 5. Time spent by Jaytee at the window on evenings when Smart was not coming home during the experiment, in 10-minute periods. Averages from ten evenings.

In public lectures and on TV shows, Wiseman claimed over and over again that he had refuted Jaytee's abilities. As recently as April 2004, he was still making this claim on his website:—

Dr Matthew Smith (Liverpool Hope University) and Prof Wiseman conducted four experiments examining the claim that a Yorkshire terrier named Jaytee could psychically detect when his owner was returning home. The results of these experiments did not support the existence of any paranormal communication between the owner and her pet. This research was widely reported in the media and published in *The British Journal of Psychology*.

In response, Sheldrake claims ([www.sheldrake.org](http://www.sheldrake.org)) "his presentations are deliberately misleading."

He makes no mention of the fact that Jaytee waits by the window far more when Pam is on her way home, nor does he refer to my own experiments. He gives the impression that my evidence is based on one experiment filmed by a TV company, rather than on more than two hundred experiments, and he implies that he has done the only rigorous scientific tests of this dog's abilities. I confess that I am amazed by his persistence in this deception.

Here, then, we have a case in which Wiseman replicated a successful psi experiment, and then attempted to explain away his successful replication by arbitrarily ignoring most of his own data.

*Meta-Analyses and Retrospective Data Selection*

Wiseman began his *Skeptical Inquirer* article by stating his central point, that “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” (p.36). This, however, is committing the fallacy of confusing *absence of evidence* with *evidence of absence*. The fact that we fail to observe positive results for a phenomenon in any individual experiment does not count as evidence that the phenomenon in question does not occur. Individual experiments may fail to show positive results for any number of reasons: the experiment may not have been performed properly; the sample size chosen may have been too small to reveal statistically-significant effects; and so on. With psi we have the added complication that we are dealing with a purported human ability, and few human abilities are perfectly replicable on demand. To use a baseball analogy, home runs are not perfectly replicable on demand, but that does not mean that home runs do not happen. And our failure to observe even a single home run at an individual baseball game does *not* count as evidence that home runs do not happen. Similarly, our failure to find evidence of psi in any individual experiment is not “evidence that psi does not exist.”

Before we come to that conclusion, we must consider *the data as a whole*. In practice, this means employing the widely-used statistical technique of meta-analysis, in which the data from several experiments of the same type are combined and then analysed as a whole. In fact, Richard Wiseman is familiar with this technique, and has used it himself to conduct a meta-analysis of the results from thirty ganzfeld psi experiments. He mentions this study on page 38:—

In 1999 Milton and Wiseman published a meta-analysis of all ganzfeld studies that were begun after 1987 and published at the start of 1997, and they noted that the cumulative effect we both small and non-significant.

But what Wiseman does not mention is that it later turned out that Milton and Wiseman had botched their statistical analysis of the ganzfeld experiments by failing to consider sample size. Dean Radin simply added up the total number of hits and trials conducted in those thirty studies (the statistically-correct method of doing meta-analysis) and found a statistically significant result with odds against chance of about 20 to 1 (Radin, 2007, pp.118, 316).

The 30 studies that Milton and Wiseman considered ranged in size from 4 trials to 100, but they used a statistical method that simply ignored sample size ( $N$ ). For instance, say we have 3 studies, two with  $N = 8$ , each giving 2 hits (25%), and a third with  $N = 60$ , giving 21 hits (35%). If we ignore sample size, then the unweighted average percentage of hits is only 28%; but the *combined* average of all the hits is just under 33%. This, in simplest terms, is the mistake they made. Had they simply added up the hits and misses and then performed a simple one-tailed  $t$ -test, they would have found results significant at the 5% level. Had they performed the exact binomial test, the results would have been significant at less than the 4% level, with odds against chance of 26 to 1. Statistician Jessica Utts pointed this out at a meeting Dean Radin held in Vancouver in 2007, in which he invited parapsychologists and sceptics to come together and present to other interested (invited) scientists. Richard Wiseman

was present at this meeting, and was unable to offer any rational justification for his botched statistics. Nevertheless, Wiseman mentions this study of his in his *Skeptical Inquirer* article, writing that “the cumulative effect was small and insignificant.”

And this was not the only problem with the study. Milton and Wiseman did not include a large and highly successful study by Kathy Dalton (1997) on account of an arbitrary cut-off date, even though it was published almost two years before Milton and Wiseman’s paper; had been widely discussed among parapsychologists; was part of a doctoral dissertation at Julie Milton’s university; and was presented at a conference *chaired by Wiseman* two years before Milton and Wiseman published their paper.

Here we have a case in which Wiseman nullified a positive result by first engaging in “retrospective data selection” — arbitrarily excluding a highly successful study — and then by mishandling the statistical analysis of the remaining data.

#### CONCLUSIONS

Here we have three clear-cut cases in which Wiseman adopted a “heads I win, tails you lose” strategy: that is, using tricks to ensure he gets the results he wants to present.

How does he do it? He has two basic techniques:—

1. Ignoring statistical methods and standards that are commonly accepted in all areas of scientific inquiry.
2. Arbitrarily excluding data that run counter to his *a priori* opposition to the existence of psi.

What can be done about this? Simply being aware of Wiseman’s history of using these tricks to dismiss positive results helps ensure that we will check to see if either technique has been used whenever he appears in the media to debunk the work of professional researchers. Also, as I have argued at length elsewhere (Carter, 2010), we need to remember always that many controversies in science have a strong ideological component, and so what is presented as good science is occasionally—upon closer inspection—nothing of the sort. And finally, regarding Wiseman’s assertion that there is no consensus regarding the existence of psi, we need to keep in mind the words of physicist Max Planck, (1950, pp.33–34) one of the founding fathers of quantum mechanics, who sadly remarked in his autobiography:—

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.

2 Halkirk Bay  
Winnipeg  
CANADA R2K 2V7

CHRIS CARTER

#### REFERENCES

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

- Blackmore, S. (1999) If the truth is out there, we've not found it yet. *The Times Higher Education Supplement*, 27 August 1999, 18.
- Carter, C. (2007) *Parapsychology and the Skeptics*. Pittsburgh, PA: Sterlinghouse Books.
- Carter, C. (2010) Persistent denial: a century of denying the evidence. In Krippner, S. and Friedman, H. (eds.) *Debating Psychic Experience: Human Potential or Human Illusion?* Santa Barbara: Praeger.
- Dalton, K. (1997) Exploring the links: creativity and psi in the ganzfeld. *Proceedings of Presented Papers. The Parapsychological Association 40th Annual Convention*, 119–134.
- Hyman, R. and Honorton, C. (1986) A joint communiqué: the psi ganzfeld controversy. *JP* 50, 351–364.
- Milton, J. and Wiseman, R.. (1999) Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin* 125 (4), 387–391.
- Planck, M. (1950) *Scientific Autobiography*. London: Williams & Norgate.
- Radin, D. (1997) *The Conscious Universe: The Scientific Truth of Psychic Phenomena*. San Francisco: HarperCollins.
- Radin, D. (2006) *Entangled Minds*. New York: Simon & Schuster.
- Sheldrake, R. (1999a) *Dogs That Know When Their Owners are Coming Home*. New York: Crown Publishers.
- Sheldrake, R. (1999b) Commentary on a paper by Wiseman, Smith and Milton on the 'psychic pet' phenomenon. *JSPR* 63, 306–311.
- Sheldrake, R. (2000) A dog that seems to know when his owner is coming home: videotape experiments and observations. *Journal of Scientific Exploration* 14, 233–255.
- Wiseman, R. and Schlitz, M. (1997) Experimenter effects and the remote detection of staring. *JP* 61, 197–207.
- Wiseman, R., Smith, M. and Milton, J. (1998) Can animals detect when their owners are returning home? An experimental test of the 'psychic pet' phenomenon. *British Journal of Psychology* 89, 453–462.
- Wiseman, R., Smith, M. and Milton, J. (2000) The 'psychic pet' phenomenon: a reply to Rupert Sheldrake. *JSPR* 64, 46–49.
- Wiseman, R. (2010) 'Heads I win, tails you lose': How parapsychologists nullify null results. *Skeptical Inquirer* 34 (1), 36–39.
- Zingrone, N. (2002) Correspondence. *JP* 66, 209–210.