Fairy tales or horror stories?   
A reconsideration of the Blackmore-Sargent controversy

I have devoted quite a few of these articles to scrutinising the claims of sceptics to see whether their criticisms should be taken seriously, and if so what consequences they have for how we conduct and interpret our research. I was interested, then, to read David Marks’ new book *Psychology and the Paranormal*. Marks devotes a chapter to experiments that adopt the ganzfeld technique,[[1]](#endnote-1) which is encouraging, since ganzfeld research still provides one of the strongest lines of evidence for psi effects in the laboratory.[[2]](#endnote-2) However, he spends little time on the extensive database of 113 studies comprising nearly 5,000 trials that have been carried out since 1974,[[3]](#endnote-3) choosing instead to focus on a controversy concerning Sue Blackmore and Carl Sargent. He acknowledges that ‘four of the five meta analyses … reported statistically significant … results, which can be interpreted as positive and replicable evidence of psi’, but sets this aside because ‘the waters have been muddied, or poisoned even, by accusations of data manipulation and fraud. These same four meta-analyses yielding positive psi hit rates included the highly contentious studies conducted by Carl Sargent’ (pp. 130-131). In fact, two of those positive meta analyses, by Bem and Honorton (1994) and by Storm, Tressoldi and Di Rissio (2010) did not include any of Sargent’s work, so that Marks’s claim that overall effects are inflated by these questionable experiments is demonstrably false. Nevertheless, it is important to be clear on whether ‘serious doubt has been cast’ on these studies, as Marks claims, and thus whether they should be excluded from future summary evaluations. I have considered the case in some detail in an essay for the Psi Encyclopedia, which I’ll summarise here.

The ‘serious doubt’ that Marks refers to derives from a report by Blackmore following a visit to Sargent’s Cambridge laboratory in 1979. After failing to produce evidence for psi in a ganzfeld experiment that formed part of her PhD, she had hoped to observe at close quarters a colleague who was using the same methods but with much greater success. There is an art to conducting research in the social sciences, and the kind of tacit knowledge that can facilitate success is, by definition, difficult to articulate and rarely features in published reports, so that visits of this type can be extremely informative. Blackmore stayed for 8 days and witnessed 13 sessions, including 6 hits, well above the 25% expected by chance. She submitted a report to the SPR concerning the visit, describing the experimental approach but also raising concerns about discrepancies she had observed that she interpreted as evidence of cheating by Sargent. She shared her report with Sargent and encouraged him to produce a response, but when it became clear that he would not, she published a slightly emended version in the July 1987 issue of JSPR.

In the 1980s many of the features of the ganzfeld experiment (such as using a random method to choose the target image, providing the participant with a set of images for judging, and recording the outcome) had to be managed by the researchers themselves. Such studies came to be labelled ‘manual ganzfeld’ to distinguish them from later ‘autoganzfeld’ designs that could employ computers to automatically manage those elements. Even with the best of intentions, there is much greater scope for human error with a manual ganzfeld design, but also greater opportunities for fraud.

Blackmore’s concerns focused on the method by which pictures were selected as targets. Sargent’s team used a database of 108 pictures arranged in 27 sets of 4. There were two copies of each picture: one was placed in an individual sealed envelope to be provided to the agent as the target; one was included in an envelope with the other pictures of its set to be used by the subject during judging. This is an important feature, since it ensures that there will be no handling marks on the picture the agent has been concentrating on that might indicate to the subject that it was the target. The target needs to be selected by some random method that ensures the subject (and indeed their experimenter) can’t second guess what it might be (e.g. by avoiding pictures with children because yesterday’s target was a children’s party). A researcher who does not have contact with the subject would consult published tables of random numbers and the first digit that fell in the range 1-28 (13 wasn’t used) determined which picture set to use for that trial. Which image within that set would be the target was decided by taking a pile of 20 small envelopes (5 containing As, 5 with Bs, etc.), cutting the deck and then counting down to the nth envelope, where n was the first number between 1 and 20 in the table of random numbers.[[4]](#endnote-4) The agent would take the four large envelopes that belonged to the set for that trial, along with the small envelope that would tell them which of the large envelopes, A-D, to open. They only opened the small envelope once they were secure in the agent’s room and the trial had begun, and would then open the large envelope indicated to reveal just that picture. Meanwhile, the participant would experience ganzfeld stimulation and give an ongoing commentary of their experience while being monitored by the session experimenter. Once the ganzfeld period was over, the experimenter would go to the office to retrieve the judging envelope that had been left out for them and present the participant with all four images that it contained for comparison against their impressions. After the trial, the small envelope wasn’t returned to the deck; instead, a replacement envelope was added. These envelopes were already sealed and unmarked so were kept in separate drawers so that the correct letter could be added to ensure there were equal numbers of each for the next trial (i.e., if the target for that trial had been a ‘C’ then an envelope from the ‘C’ drawer would be added to the deck).

Blackmore speculated that the deck of small envelopes might be biased rather than having equal numbers of the four options, A-D, so that some targets would be more likely than others. She also wondered if the target letter for a particular session might be taken from one of the drawers rather than from the shuffled deck, so that its identity would be known with certainty. She focused on trial 9 in the sequence she observed after discovering that one of the envelopes in the B drawer was missing, and the target for that trial had been picture B. Sargent was not formally involved in the trial, but had stepped in to act as randomiser and was then present during the judging phase and ‘seemed to push the subject towards picture B’, which would be completely inappropriate if he had some way of knowing which letter had been chosen as the target in this set. The subject ranked picture ‘B’ as their selection, and this did turn out to be the target image, so the trial was recorded as a ‘hit’.

Blackmore disclosed her concerns to Trevor Harley, Sargent’s co-experimenter, and they checked all of the envelopes in the deck to see if each letter was equally represented, as they should be. However, they discovered an extra ‘A’ and ‘B’ and one fewer ‘C’ and ‘D’ than expected. While clearly this deviation from expectation should not have happened at all, it is much too small to have any substantial effect on the study outcome, and so provides only meagre evidence in support of Blackmore’s first mechanism. If the mechanism for cheating was to take a small envelope directly from one of the drawers (or from elsewhere in the room) to give to the agent so that the target image was known for certain, then it’s not clear why there would be any need for the distribution of letters in the deck to be biased. Harley and Blackmore also conducted a search of the office, looking for a lab book that was used to record details of each session. The book was not found, but they did discover some additional small envelopes in various places (under papers, in another drawer). These were sealed, and when opened revealed a ‘C’, a ‘D’ and a batch of 3 ‘A’s. Again, this is unexpected, but their haphazard placement is very difficult to fathom, or to align with the other observations - if known letters were secreted about the office, then why should drawer numbers fall unexpectedly, and why should the shuffled deck have uneven numbers? It would have been much more straightforward to simply have spare cards and open envelopes and then write whichever letter was needed and seal it in the envelope; this would have taken no more time than hunting around for whichever secret location contains the ‘correct’ envelopes. For a supposedly astute investigator, the discovered method seems very naïve indeed.

Blackmore went on to make some predictions based on her suspicions, most of which sadly remain untested. However, she did postulate ‘if one person were cheating, the most significant results should occur when they were acting as agent or experimenter’, and claimed that indeed scores were higher in sessions that involved Sargent in one of these roles. However, my own analysis of the data to which Blackmore refers (described in detail in the encyclopedia entry) actually shows that they are contrary to her predictions, and in one case relies on a misunderstanding of the data presented. Similarly, while we may be concerned to learn of Sargent ‘pushing’ the subject towards a picture that turned out to be the target in one trial, that concern may be allayed when we learn of another trial in which Sargent urged the subject to reconsider correspondences to one of the pictures that turned out not to be the target (the subject was unconvinced and stuck with their own preference, which was the target). If Sargent’s ‘pushing’ had been successful, it would have prevented a hit rather than created one. More generally, the efficacy of this kind of subtle social pressure has to be evaluated in the context of striking qualitative correspondences between ganzfeld imagery and target features (reported, for example, by Ashton et al.).

It is clear from this brief review that Blackmore’s concerns about Sargent’s research practices are technical and quite complex, and the proposed mechanisms for cheating rely on the convergence and interplay of rather subtle factors (e.g., in purposely choosing a picture that is expected to align with the subject’s personal preferences, or knowing the target choice and relying on experimenter pressure during judging to override overt similarities between the subject’s ganzfeld impressions and the imagery found in the 4 pictures). To an extent, this complexity and subtlety is appreciated in the exchange that was published in the SPR’s Journal in 1987. Unfortunately, more recent accounts of the controversy tend to be much briefer and more polemical, as we saw with David Marks’s comments at the beginning of this article. Similarly, when Blackmore revisited the case in a 2017 *Skeptical Inquirer* article,[[5]](#endnote-5) she wrote ‘it became clear that Sargent had deliberately violated his own protocols and in one trial had almost certainly cheated.’ In my judgement, such a definite conclusion is not justified by the material I’ve just reviewed.

Blackmore goes on, ‘it matters that Sargent’s experiments were seriously flawed. It matters that Bem included [them] in his meta-analysis without referencing the doubt cast on them. It matters because Bem’s continued claims mislead a willing public into believing that there is reputable scientific evidence for ESP in the ganzfeld when there is not.’ It is quite astonishing for Blackmore to extrapolate from her suspicions about practices she observed at one laboratory during an 8-day visit so as to dismiss 35 years of research carried out by 46 different principal investigators, about 70% of which are of the ‘autoganzfeld’ type to which concerns about randomisation and target selection cannot, by definition, apply. Recent experiments show no indication of a decline,[[6]](#endnote-6) and are not dependent on the success of any particular researcher, so that a meta-analysis that excluded Sargent’s work would still be highly significant. However, on the basis of the material reviewed here, there are absolutely no grounds for doing so.

1. If readers are unfamiliar with this technique, I recommend the Psi Encyclopedia’s essay at <https://psi-encyclopedia.spr.ac.uk/articles/ganzfeld-esp> [↑](#endnote-ref-1)
2. Lance Storm & Patrizio Tressoldi, Meta-analysis of free-response studies 2009-2018: Assessing the noise-reduction model ten years on. Journal of the Society for Psychical Research, 84 (2020), 193-219. [↑](#endnote-ref-2)
3. Patrizio Tressoldi & Lance Storm, Stage 2 Registered Report: Anomalous perception in a ganzfeld condition - A meta-analysis of more than 40 years investigation. (2021) <https://f1000research.com/articles/10-234> [↑](#endnote-ref-3)
4. This description is from Hugh Ashton, Peter Dear, Trevor Harley, & Carl Sargent, a four-subject study of psi in the ganzfeld. Journal of the Society for Psychical Research, 51 (1981), 12-21. Blackmore’s description is more convoluted, but may be based on her personal observations. [↑](#endnote-ref-4)
5. Sue Blackmore, Daryl Bem and psi in the ganzfield [sic]. *Skeptical Inquirer*, *42*(1) (Jan/Feb 2018), available at <https://skepticalinquirer.org/2018/01/daryl-bem-and-psi-in-the-ganzfield/> [↑](#endnote-ref-5)
6. Tressoldi & Storm, op cit. [↑](#endnote-ref-6)