# A Critical Response to Rick Berger

#### SUSAN J. BLACKMORE

ABSTRACT: Berger set out to do a meta-analysis of my ESP experiments but ended up doing nothing constructive at all. He confined himself to the ESP experiments 1 carried out over 10 years ago (between 1976 and 1978) for my doctoral dissertation. I have responded to his criticisms section by section, noting where he has made errors in dating experiments, has confused one experiment with another, has accused me of omitting a disclaimer from a published paper when it is clearly included, or has made other unfounded accusations. In these and other ways he has seriously misrepresented my work. However, in spite of Berger's numerous errors, I agree that one cannot draw conclusions about the reality of psi based on these experiments. The results are relevant to the problem of replicability in parapsychology, but as far as the reality of psi is concerned I draw only one conclusion: I don't know.

Rick Berger's (1989) argument seems to be as follows: (a) There are many inconsistencies in my experimental reports, (b) the vast majority of my studies were flawed, and (c) therefore no conclusions should be drawn from my results.

I would like to respond to (a) by commenting on the "inconsistencies" (below) and to (b) by agreeing. Most of my early ESP studies had some weaknesses, and as Berger acknowledges, 1 pointed these out myself. In regard to (c), I would like to ask what conclusions he supposes I did draw. I concluded that 1 could not test my hypotheses because I found no psi, and my most general conclusion (e.g., Blackmore, 1985, 1987a) was that the evidence only justified having an open mind but that that was extremely difficult. This is a conclusion I stand by.

I should add that Berger has confined himself to criticizing the ESP experiments that I carried out more than 10 years ago (between 1976 and 1978) for my Ph.D.

When I first read Berger's criticisms, after supplying him with additional information through correspondence, I realized that he had made numerous errors and false accusations. I pointed these out, and naturally he withdrew misquotes and corrected some errors before his critical review was published. However, I would now like to respond to those that remain. I shall do this section by section in accordance with his paper.

#### THE DATABASE BROADLY VIEWED

Berger begins by writing: "Blackmore reported 29 experiments conducted between October 18, 1976 and December, 1978 ..., of which 21

> The Journal of the American Society for Psychical Research Vol. 83, April 1989

were eventually published as separate experiments in five peer-refereed parapsychology journal papers" (Berger, 1989, p. 125). I do not know how he arrived at the number 21, but the figure is incorrect. In fact, 27 of these experiments were published. Berger appears to have ignored the questionnaire, which was published in Blackmore, 1982b, with analyses showing relationships between ESP scores and the reporting of out-ofbody experiences (OBEs) and lucid dreams. In fact, only two experiments were not published. The first was a small informal test to see whether a technique that had previously failed with a clairvoyance task might be more effective with GESP (based on the Main Experiment in Blackmore, 1981b). As I explained in my dissertation (Blackmore, 1980b), there were only 9 pairs of subjects, and they were unsupervised. For this reason I did not publish the experiment. The second unpublished experiment was a Ganzfeld experiment carried out in two parts, with a total of 36 sessions. The flaws in the technique (e.g., shuffling the targets for the first 20 sessions) were detailed in the dissertation and were the main reason for not publishing the experiment.

Because Berger seems to be accusing me of covering up any positive results I might have found, I should add that neither study provided any. For the first, scores were insignificantly below mean chance expectation (MCE) (mean = 5.0, MCE = 6.0, t = 1.0, 8 df, p = .35), and there was no effect of target type (F[2,24] = 0.34). For the Ganzfeld study, the sum of ranks was at chance (50) for the first 20 sessions and insignificantly below chance (35, MCE = 40) for the remaining 16 sessions. The first 20 sessions comprised a comparison of 10 subjects trained for 6 months in imagery, relaxation, and OBE induction with 10 untrained subjects. The trained subjects showed increased scores on a shortened form of Belts' Questionnaire Upon Mental Imagery (QMI) (vividness of imagery scale), but they did not score higher in a clairvoyance test with ESP cards nor in the Ganzfeld ESP task. There was no difference in imagery scores between hitters and missers in the Ganzfeld. Further details are given in the dissertation (Blackmore, 1980b). These two are the only experiments from the dissertation that were not published.

1 suspect that Berger made this error because he did not take into account that many experiments consisted of two or three pans. For example, to look for correlations between ESP and memory, I might have given an ESP test one week, a memory test the next, and a further ESP test in the final week. Also, data from one test might be used in more than one experiment: for example, to correlate the same ESP scores with two different kinds of memory test. In the schedule of experiments to which Berger often refers, all sessions are coded and dated to make this clear. It may be a complex task to identify the details of each experiment, but it certainly is possible because 1 have done it (10 years later) in preparing this response. It was to make this precise reconstruction possible that I gave so many details and included the schedule in the dissertation in the first place. It provides a much more detailed account of what was actually done than is typical of most published scientific work. Berger's error (of claiming 21 instead of 27 of the experiments published) shows that he has not been able to make use of this information accurately.

In a subsection entitled " 'Ten Years' of Negative Research," Berger states: "The primary implication of Blackmore's recent skeptical publications is that her 'ten years of negative research' (see Blackmore, 1987[a]) is a sound basis upon which she may conclude and promote the notion that parapsychology should be redefined as 'a new psychical research—one without psi' " (p. 127). But I have never concluded that there are no psi phenomena. I have only concluded that I cannot find them, and I have wondered (perhaps *ad nauseam*) what that tells either me or anyone else.

Berger is right that in an address to the 1986 conference of the Committee for the Investigation of Claims of the Paranormal (CSICOP), I described how my mind seemed to change "from closed belief to closed disbelief" (Blackmore, 1987a, p. 249), but he does not explain that I was trying to compare the problems of cognitive dissonance that beset both the extreme believer and the extreme disbeliever. He simply stops the quote there. In fact, I said next: "But either way J suffered," and I went on to discuss the evidence that the disbeliever has to take into account:

I am thinking particularly of the results of Carl Sargent, Charles Honorton, Helmut Schmidt, and Robert Jahn. I suggest that if we think these can easily be dismissed then we are only deluding ourselves. One cannot offer simplistic counterexplanations and throw all these results away. (p. 250)

In other words, 1 was explaining precisely why one *cannot* conclude that there are no psi phenomena.

# OVERVIEW OF THE DATABASE

In this section, Berger lists some of my published papers and the problems he claims to have identified in them. Most are repeated in later sections, but one or two are not mentioned again so I shall deal with them here.

Berger's implication seems to be that I noted flaws in my dissertation and then covered them up in published papers. This would, if well founded, be a rather serious accusation. However, it is not well founded. On page 129 he quotes my dissertation: "There were too few trials to conclude that there is no effect" and that "the results of this exploratorystudy are included only for the sake of completeness" (Blackmore, 1980b, p. 171). He then writes: "In the published version (Blackmore, 1981a, "Target"), no such disclaimer is noted." Bearing in mind that a published paper must be briefer than a dissertation, it is quite appropriate that the published version actually states: "The results obtained do not confirm any of the hypotheses but there was only one subject and few trials. A second experiment was therefore carried out with more subjects" (Blackmore, 1981a, p. 11). Berger does, however, point out two genuine errors. I did report the number of subjects as 23 instead of 28 in Experiment 4 (I assume this is a typographical error. The degrees of freedom are correctly given as 27). Also, a *p* value of .52 was erroneously given instead of .052 in Experiment 1 (Blackmore, 1981a). The value of *t* is correctly given (t = 2.74), thus the error should be obvious to other readers as well.

In Berger's discussion of my Ganzfeld experiment, to answer his implied questions: Yes, of course it was included in my survey of unpublished Ganzfeld studies, and naturally it was counted as flawed. It was, after all, seriously flawed, and this is why I did not publish it. Had it provided significant results, 1 would certainly have gone on to carry out better controlled Ganzfeld studies.

# SPECIFIC CRITICISMS

In a new section, Berger's first criticism is that I reordered the experiments in my publications. I presume he means more than that I reported experiments in the most logical way irrespective of the precise»order in which they were completed. I presume he is implying that in some sense I sought to mislead or confuse my readers. In any case, he has the facts wrong.

For his first example, regarding the six experiments on correlations between ESP and memory reported in Blackmore, 1980a, he claims that the order was 3, 1, 4, 2, 6, 5 (Berger, 1989, p. 134). This plus the repeated error of claiming that 21 (not 27) experiments were published makes my response difficult, but I shall try to clarify the correct order. Any experiment in the text of the dissertation or in a published paper may have appeared as two or three sessions in the "Schedule of Experiments," and some experiments overlapped with others. To describe the order in which a series of experiments was done, one must take into account both starting and finishing dates. In his Table 2, Berger has mixed up starting, finishing, and intermediate session dates. If one takes the starting dates for these experiments, the actual order is 1 & 3, 2 & 4, 5, 6, which is perfectly logical. Experiments 1 and 2 were similar and thus were reported together, as were 3.4, and 5. Experiments 1 & 3 and 2 & 4 used data from the same initial tests, and of course their starting dates are the same. If one takes the completion dates, the order becomes 3, 1,4, 2, 6, 5. It seems to me that the only issue of any importance here is whether, as I claimed, the final experiment could be said to have been based on the previous ones. Berger writes: "The 'final' experiment (completed, according to the dissertation chronology, on December 4, 1978) preceded the fifth experiment (completed December 11, 1978) by one week" (p. 134). In fact, it preceded by a week the final part of the final experiment in a series of three similar experiments, each of which had two or three parts. I was therefore able to take account of the problems of this and the previous experiments when I designed Experiment 6.

In Berger's second example, Regarding Blackmore, 1981 a, Experiments 1 and 2 were performed after the others. This paper details three separate kinds of experiments on the effect of target types on ESP scores. Experiments 1 and 2 used stimuli of varying information content. Experiments 3 and 4 tested whether or not an agent learning the pairs of words made any difference. Experiments 5 and 6 explored target memorability. I did note that "problems found in the previous experiments were eliminated" (Blackmore, 1981a, p. 19), but this sentence goes on: "and all the subjects had individual target orders.", I was clearly referring to the specific problem of a possible stacking effect—a problem noted in Experiments 3, 4, and 5 but absent from 1 and 2. Berger does point out that reordering the experiments reveals a (nonsignificant) "substantial decline" (p. 135) in the ESP scores. It would be interesting to see whether this applies to the whole database.

Berger goes on to suggest that I invoked study quality when the outcome was significant and that I ignored it when it was nonsignificant. I find this most difficult to respond to. As far as I could, I detailed all flaws I considered to be of any importance, and indeed, many might better be considered "less than optimal design" rather than "flaws." My negative conclusions were not usually based on the flaws at all but on the fact that the results of series of experiments never seemed to produce consistent results.

A possible exception is the Tarot experiments (Blackmore, 1983b). It now appears from Markwick's (1988) analysis that the first of the three experiments does remain significant after a new analysis. Had I had the results of this reanalysis 10 years ago, I would have undoubtedly gone on to try to find out why the later experiments did not work (looking perhaps at subject variables, changes in the relationships between the subject and experimenter, and so on). As it was, I made the reasonable assumption, given what I knew, that the flaw was responsible for the original effect. I should also add that the experiments showed how face-to-face Tarot readings can be so effective, and this contribution remains, regardless of whether the "test readings" showed any paranormal effect or not.

I think the crux of this argument is shown in another quote Berger attributes to me, in which I wrote: "These faults, however, might be expected to produce spurious differences, but are unlikely to be responsible for the uniformly chance results obtained here" (Blackmore, 1980b, p. 181). I stand by this conclusion.

The importance of this argument is that Berger suggests that I am using different standards for negative outcomes than for positive outcomes and that I overlook my own flaws and view the database as "a coherent body of evidence that converges on the conclusion that psi does not exist" (Berger, 1989, p. 137). But this is unfair. I have been at pains to explain how difficult it is to draw conclusions from these negative results. He also describes "the paradox exemplified by the Blackmore work: Had such work produced consistently positive outcomes, the results could all be

dismissed as having arisen from design flaws" (Berger, 1989, p. 137). But this misses the point entirely—I would not have done those particular experiments in that way had I obtained any signs of ESP. I originally hoped to test theories about the relationship between ESP and memory, and I carried out preliminary experiments. Had these been successful, I would have gone on to improve the experiments, eliminating any flaws that might give spurious results and exploring the relationship further. As it was, I failed to find any evidence for psi and instead tried various methods in hopes of finding an experimental paradigm with which to test the theories. In this 1 failed. The kinds of flaws and their importance must be seen in this context.

Of course, there are quite different kinds of flaws that would make any negative results even less informative. Berger hints at the type of flaw that can produce spurious negative outcomes, namely Type II errors, though he says he is leaving this to a future paper. The most obvious of these is sample size. On page 129 he discusses my experiments with children as follows: "Whereas Spinelli had tested 1,000 subjects to achieve his reported results (Spinelli, 1977), Blackmore used 19 and 48 children in her two studies. Neither of Blackmore's studies showed an overall psi effect."

His implication seems to be that I had far too few subjects to detect the effect were it present. Indeed, Berger previously criticized me in an earlier version of his paper for not providing an estimate of Spinelli's effect size to allow comparison of the statistical power of my experiments to detect effects such as those Spinelli reported. In fact, it is impossible to do an accurate analysis of Spinelli's results. I have argued elsewhere (Blackmore, 1984a) that Spinelli's major experiments cannot be counted as evidence for psi because the sender could choose the target. Only his later studies, with 200 subjects, used preselected targets. Nevertheless, let us ignore this for a moment and assume that all Spinelli's work gives a true estimate of the effect size for ESP in young children. The question is then: Did I use a large enough sample to be able to detect the effect?

The age group in question is 3-5 years. This spans Spinelli's first two age groups. For each he provides three remarkably similar estimates of success rates—or percentage hits (MCE is 20%). For age 3.3-3.7 he gives 45.1, 44.0, 45.2, and for ages 4.5-4.9 he gives 36.1, 33.4, 34.8. On this basis, we may assume that the mean percentage of hits is 39.8-or roughly 40%, double the hit rate expected by chance.

To make an estimate of the required sample size to detect this effect one needs a measure of sample error or variance. However, in his paper Spinelli (1977) gives only p values and does not even say what analysis was used. In his doctoral dissertation (Spinelli, 1978), it is clear that his main analysis is a chi square—but an invalid one pooling all trials for all subjects. He gives no estimates of error or variance, which makes it impossible to estimate the sample size required.

The problem is that we do not know from Spinelli's faulty analysis whether the effect was attributable to the performance of a few good subjects or was uniformly distributed. We could, however, consider the extremes. On the one hand, if a few subjects got perfect scores (10/10) and the rest scored at chance (2/10), to obtain 40% hits a whole 25% of subjects would have had to be scoring perfectly. One would not need many subjects to find some like this. On the other hand, if all subjects were contributing equally to the effect, then one can pool all trials and use a normal approximation to the binomial to estimate the standard deviation. I can then calculate the number of subjects required to obtain a z > 2.58 (i.e., an outcome significant at p < .01). This gives N = 40. In other words, I would have needed 40 trials to detect the effect—or just 4 subjects. Testing 48 subjects, the chances of my failing to detect a 40% hit rate would be  $p < 10^{-7}$ . Of course, one might not expect to get an effect as large as Spinelli's with a different experimenter, different subjects, and so on. Therefore, it is useful to increase sample size to try to detect a weaker effect. Even allowing for this, it is clear that I used a perfectly adequate sample.

Apart from weak or inappropriate statistical tests (which I have already discussed in the context of the Tarot experiments), this is the only example for which Berger gives any details of a flaw likely to produce a Type II error. He is apparently planning at a later time to deal with others, and he mentions such problems as sampling from inappropriate populations, experimenter expectancy effects, demand characteristics, and the faulty operationalization of dependent measures. Most of these are discussed in my publications, and indeed the progression of my experiments can be seen as an attempt to deal with them. I look forward to seeing what further analyses he brings to bear.

In the section entitled "Misreporting the Original Data," Berger's first example is correct. I did indeed unintentionally give an erroneous impression of where the idea of two significant outcomes in 34 tests came from. I was previously unaware of this and am grateful to Berger for pointing it out. In his second example, however, he is guite wrong. This concerns the experiment on errors described in my autobiography. He accuses me of reporting a positive and significant outcome in my dissertation and published paper and then pretending it is negative in my autobiography (Blackmore, 1986a). He also claims that "Blackmore seems to be arguing that a flawed study with a significant outcome is equal to a negative outcome" (p. 139). In fact, I do nothing of the kind. Berger's error is that he has confused two experiments. The one reported in my autobiography as "one of my first experiments" (Blackmore, 1986a, p. 34) is not Pilot Study 1 (Blackmore, 1981b, Experiment 1, which did obtain significant differences but had a possible stacking effect). It was the second pilot study (which used an improved method, produced no significant effects, and had no sign of the pattern of results found in the first experiment). It should have been obvious to Berger which experiment was intended because they use different methods. In the first, subjects tried to draw the target, whereas in the second they chose one from a set of pictures. These methods are clearly described in both the book and the paper (Blackmore, 1981b, pp. 55, 58; 1986a, p. 34).

Berger concludes that "the claim of 'ten years of psi research' actually represents a series of hastily constructed, executed, and reported studies that were primarily conducted during a 2-year period" (p. 140). This also is unfair. It is he who has restricted the 10 years to experimental "psi research," not I. Almost all of these early experiments were done during the process of learning to do research for a doctoral dissertation, and I hope it can be seen that the methods used improved. There is much more to my research in parapsychology than this series of psi experiments. There are many years of research on OBEs (see, e.g., Blackmore, 1982a, 1982b, 1983a, 1984b, 1984c, 1986b, 1986c, 1987b) and lucid dreams (Blackmore, 1988a), experiments on probability and belief in psi and the illusion of control (Blackmore & Troscianko, 1985), and recent work on near-death experiences and the tunnel experience (Blackmore, 1988b). Based on this work, it seems to me that the important, and even lifechanging, aspects of these experiences can be better understood not by invoking psi but by trying to understand the nature of changes in consciousness. About psi I have no "strong convictions." I have done few psi experiments in recent years because I no longer expect them (for me at least) to provide positive results. I prefer to leave them to others, like Rick Berger, who take a different view and may yet prove me wrong.

## CONCLUSIONS

Berger set out to do a meta-analysis of my ESP experiments but has ended up doing nothing constructive at all. He accuses me of "distorted information," but in return he has made numerous errors and has seriously misrepresented my work. Nevertheless, I am glad to be able to agree with his final conclusion—"that drawing *any* conclusions, positive or negative, about the reality of psi that are based on the Blackmore psi experiments must be considered unwarranted" (Berger, 1989, p. 141). As far as the "reality of psi" is concerned, I can draw only one conclusion. It is one I have often expressed before and with which I ended my autobiography (Blackmore, 1986a), and that is simply, "I don't know."

### REFERENCES

- BERGER, R. E. (1989). A critical examination of the Blackmore psi experiments. *Journal of the American Society for Psychical Research*, 83, 123-144.
- BLACKMORE, S. J. (1980a). Correlations between ESP and memory. *European Journal of Parapsychology*, *3*, 127-147.
- BLACKMORE, S. J. (1980b). Extrasensory Perception as a Cognitive Process. Unpublished doctoral dissertation, University of Surrey, Guildford, England.

- BLACKMORE, S. J. (1981a). The effect of variations in target material on ESP and memory. *Research Letter*, No. 11, 1-26.
- BLACKMORE, S. J. (1981b). Errors and confusions in ESP. European Journal of Parapsychology, 4, 49-70.
- BLACKMORE, S. J. (1982a). Beyond the Body: An Investigation of Out-ofthe-Body Experiences. London: Heinemann.
- BLACKMORE, S. J. (1982b). Out-of-body experiences, lucid dreams, and imagery: Two surveys. *Journal of the American Society for Psychical Research*, *76*, 301-317.
- BLACKMORE, S. J. (1983a). Birth and the OBE: An unhelpful analogy. Journal of the American Society for Psychical Research, 77, 229-238.
- BLACKMORE, S. J. (1983b). Divination with Tarot cards: An empirical study. *Journal of the Society for Psychical Research*, 52, 97-101.
- BLACKMORE, S. [J.] (1984a). ESP in young children: A critique of the Spinelli evidence. *Journal of the Society for Psychical Research*, 52, 311-315.
- BLACKMORE, S. J. (1984b). A postal survey of OBEs and other experiences. *Journal of the Society for Psychical Research*, 52, 225–244.
- BLACKMORE, S. J. (1984c). A psychological theory of the out-of-body experience. *Journal of Parapsychology*, 48, 201-218.
- BLACKMORE, S. [J.] (1985). The adventures of a psi-inhibitory experimenter. In P. Kurtz (Ed.), A Skeptic's Handbook of Parapsychology (pp. 425-448). Buffalo, NY: Prometheus Books.
- BLACKMORE, S. [J.] (1986a). *The Adventures of a Parapsychologist*. Buffalo, NY: Prometheus Books.
- BLACKMORE, S. J. (1986b). Out-of-body experiences in schizophrenia: A questionnaire survey. *Journal of Nervous and Mental Disease*, 174, 615-619.
- BLACKMORE, S. J. (1986c). Spontaneous and deliberate OBEs: A questionnaire survey. *Journal of the Society for Psychical Research*, 53, 218-224.
- BLACKMORE, S. J. (1987a). The elusive open mind: Ten years of negative research in parapsychology. *Skeptical Inquirer*, 11, 244-255.
- BLACKMORE, S. J. (1987b). Where am I? Perspectives in imagery and the out-of-body experience. *Journal of Mental Imagery*, 11(2), 53-66.
- BLACKMORE, S. J. (1988a). A theory of lucid dreams and OBEs. In J. Gackenbach & S. LaBerge (Eds.), *Conscious Mind: Sleeping Brain* (pp. 373-387). New York: Plenum.
- BLACKMORE, S. J. (1988b). Visions from the dying brain. *New Scientist*, 118(1611),43-46.
- BLACKMORE, S. [J.], & TROSCIANKO, T. (1985). Belief in the paranormal: Probability judgements, illusory control, and the "chance baseline shift." *British Journal of Psychology*, 76, 459-468.
- MARKWICK, B. (1988). Re-analysis of some free-response data. *Journal* of the Society for Psychical Research, 55, 220-222.
- SPINELLI, E. (1977). The effects of chronological age on GESP ability

[Summary]. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in Parapsychology 1976* (pp. 122-124). Metuchen, NJ: Scarecrow Press.

SPINELLJ, E. (1978). *Human Development and Paranormal Cognition*. Unpublished doctoral dissertation, University of Surrey, Guildford, England.

Department of Psychology University of Bristol 8-10 Berkeley Square Bristol BS8 1HH United Kingdom