

"Experiment One of the SAIC Remote Viewing Program: A Critical Re-Evaluation": Reply to May

Richard Wiseman and Julie Milton

Abstract

In our original paper (Wiseman & Milton, 1999), we described a number of potential information leakage pathways in Experiment One of the SAIC remote viewing program. In this paper, we counter May's assertion that the proposed leakage paths could not have accounted for the observed difference between performance in the static-target and dynamic target conditions and refute May's claim that there is empirical evidence from the literature that indicates that our proposed information pathways cannot be effective. We also described in our earlier paper the repeated and marked difficulties encountered by May and the SAIC research team in producing a consistent account of a number of aspects of the experiment's procedure, which resulted in a sequence of five different accounts of these procedures. These difficulties not only make an assessment of Experiment One extremely difficult but call into question whether the assessors commissioned to write a US government-sponsored report on the other studies in the SAIC program would have been given accurate information about their unrecorded details. In his response, May insists that the final account of Experiment One is the correct reconstruction but offers no evidence or argument to support this claim and does not address the problems that the situation raises for the government report.

Introduction

May (1998) presents a number of arguments against the existence or efficacy of the potential leakage paths in Experiment One of the SAIC remote viewing program identified in Wiseman and Milton (1998). We examine each of his arguments in turn.

1. Experimenters Altering Receivers' Responses

The first potential pathway for sensory leakage in Experiment One that we discussed concerned potential handling cues if the experimenter who selected the targets in the experiment also later handled the receivers' responses before passing them on to the independent judge. May states that ambiguous terminology in Lantz et al. (1994) misled us into mistakenly believing that the same experimenter was responsible for both the target selection and the handling of receiver responses, and asserts that the final recollection of the SAIC team that different experimenters carried out the tasks, is correct. He omits to mention that, as clearly described in our paper, the ambiguity of Lantz et al.'s account caused R.W. to email May for clarification and that May confirmed that the same individual was responsible for both tasks. The first collective memory of the SAIC team also involved having the experimenter who handled the receivers' responses being one of the three who selected the targets. May offers no arguments or evidence to support his statement that the final collective recollection is accurate. Mere eventual agreement among the SAIC team cannot be sufficient evidence in the face of their repeated difficulties in producing a consistent account of the study's procedure. The question still remains of whether their final account is reliable.

May argues, however, that two lines of evidence suggest that even if the proposed information pathway had existed, it would have been unlikely to have affected study outcome. First, he argues that if any sensory leakage had occurred in Experiment One, including that involving the information pathway under discussion here, both types of trial would have been equally affected. Because trials involving static targets outperformed dynamic-target trials to a statistically significant degree, he argues that no sensory leakage occurred in the study. May clearly believes that it would be impossible for the impact of an artifact to differ across experimental conditions. In fact, evaluators of psi research often argue that artifacts would be more effective in one condition than another and sometimes use such models to assess the likelihood of the artifact's actual existence (see, e.g., Wiseman, Smith and Kornbrot, 1996). In Experiment One it is not

difficult to imagine why the proposed leakage might only have been effective in trials involving static targets. The static targets were pictures whilst the dynamic targets comprised film clips recorded on unlabelled videotapes. Targets were sent to receivers after each trial as feedback and they then mailed them back to the laboratory. Experimenters opening the envelopes containing the returned target material would thus have seen the pictures but not the content of the videotapes. If these experimenters then came into contact with the receivers' response sheets, any inadvertent handling cues would only be placed on response sheets from the static trials.

Second, May states that it is unlikely that Experiment One could have been affected by inadvertent handling cues because, he asserts, there is "considerable evidence" that such cues do not affect the outcome of psi experiments. In fact there are only 3 published papers, all cited by May, that report studies directly examining the potential effects of sensory cues in ESP studies. May reports that Schmeidler and Lindemann (1966) demonstrated that deliberate sensory cues did not help receivers but this is clearly not the case. Although most of the receivers appeared unable to use the cues, Schmeidler and Lindemann report that 5 of the 96 receivers successfully used the cues to obtain an 89% hitting rate in 100 trials where 50% would have been expected by chance. Concerning the chance performance of the other receivers, they write: "It would be reckless to infer that other subjects tested under other conditions will also disregard incidental cues and that ESP experimenters may relax the rigor of the conditions" (p. 361). The informal studies by Susan Blackmore that May describes as also demonstrating the lack of efficacy of deliberate cues have not been published or documented. However, Blackmore confirms similar results to those of Schmeidler and Lindemann in that although overall participants scored at chance, a few participants in each study successfully detected and used the cues (personal communication to J.M.). Palmer (1983) issues a similar warning to Schmeidler and Lindemann's concerning the null results of his own study in which unselected participants failed to exploit handling cues on the target picture in a free-response "ESP" task. He points out that it is impossible to be certain to what extent his results would generalise to other studies, such as those involving selected participants (as did May's Experiment One). Finally, May notes that Palmer

and Kramer (1986) found that participants asked to look for handling cues on photographs were able to do so. This was in fact the case and indicates that at least under certain circumstances, cues of the sort that might arise naturally are available to be exploited.

Taken together, the studies do not support May's interpretation of them that handling cues do not appear to affect the outcome of ESP studies. Rather, they indicate that such cues can be effective even for unselected participants under certain conditions. Very few studies examining the effects of sensory cues have been carried out and none have involved selected participants or the specific type of handling cue being posited for Experiment One. It is therefore impossible to say whether cues of the type we propose for this study would have been effective and if so, with what effect size. However, the empirical evidence refutes the view that such cues can be assumed to be ineffective.

2. Receivers Altering Their Own Responses

In Experiment One, receivers faxed their response sheets to Lantz and then received target feedback. May clarified that receivers may have received target feedback before mailing their original response sheets to SAIC. As we point out in our paper, the receivers might therefore have placed deliberate or inadvertent target-related cues on the originals. If the "analyst" (the independent judge, May himself) used the original response sheets rather than the faxes, a pathway for sensory leakage would exist. May does not argue that the higher performance in static target trials compared to dynamic trials suggests that this particular information pathway was ineffective but it is clear that there are several possible mechanisms for such a difference. For example, receivers may have been more likely to handle their response sheets when comparing them to a picture rather than to a videotape, which would require the manual operation of a videorecorder; if the static targets were more distinct from each other than the multiple images in the video clips, any cues may have been more effective during the judging process; and so on.

May now contends that because the research team was aware of the possibility of receivers altering their original response sheets after feedback, the independent judge used the faxed version for judging. However, in his initial emails May repeatedly insisted that he used the original response sheets during judging (29 April 1996; 30 April 1996a; 30 April 1996b; 31 May 1996; 2 October 1996) and that Luke carefully compared the original and faxed sheets to ensure that they were identical. It is only in the SAIC team's final collective memory that May used the faxes. Despite the fact that the team's final recollection disagrees with his own repeated accounts of what he himself did in the experiment, he presents this account as established fact and implies that we concede its accuracy. We do not. As before, he produces no additional evidence or documentation to support this final version of events. This lack of evidence, coupled with the difficulties in producing a consistent account of the study, calls the reliability of this final account into question.

May also continues to stress that the original responses were carefully compared to the faxed version prior to judging. In his earlier accounts of this procedure (29 April 1996a; 29 April 1996b; 31 May 1996), May states that the check was performed to detect any alterations that the receivers might have made in the originals after receiving feedback because he was using the originals from which to judge. It is difficult to understand the purpose of this check if the judge was instead using faxed response sheets that had been sent by receivers prior to them obtaining feedback. The existence of this check is an additional reason to question the accuracy of the final collective memory on this point.

3. Knowledgeable Experimenters in Contact with the Analyst

In our paper we note that there was frequent contact between May (the independent judge) and the experimenters involved in target selection, raising the possibility that the experimenters may have inadvertently cued the judge. May believes that such information leakage did not occur for two reasons. First, he argues that an experimenter would have had to have known which trial the judge was analyzing - which they did not - and inadvertently leak information specific to that

trial's target in order for the information leakage to be effective. However, the mechanism for information leakage that we discuss in our paper clearly requires neither assumption. Rather, we proposed that if (as is likely by chance) particular material was being chosen unusually often as the target, the experimenters would be aware of this and the knowledge might have biased their everyday conversation towards the frequent target content without their knowledge. This content might then have been especially salient to the judge throughout all of the trials that he judged, thus introducing a target-related bias. Second, May argues that even if such information leakage had taken place it would have affected all trials equally and not just those involving static targets. Again this argument is invalid as there are several ways in which information leakage may have affected only the static target trials. For example, only static targets may have displayed slightly unusual selection frequencies; static targets may have contained material more memorable to the experimenters; the information in the static targets may have been more easily communicable by inadvertent means; because experimenters only saw the content of returned static targets rather than the unlabelled videotapes, only the static target material may have been in their minds; and so on.

May agrees that Experiment One's protocol did allow for an obvious leakage path in having contact between the independent judge and the sender, Lantz, who was aware of the identity of the targets in the sender condition. However, May notes that this potential source of leakage did not affect results as scoring was not significantly higher on the sender trials than on the no-sender trials. Here we agree with May that Lantz does not appear to have been a source of information leakage in this context. However, just because such leakage did not occur between Lantz and May under the conditions of Lantz's rare visits to SAIC, there is no reason to believe that it could not or did not occur between the other experimenters and May while working in continual and close proximity to one another at SAIC.

4. Random Judging Order

In our paper we raised the possibility that in judging the trials out of their original order, May could have been unconsciously influenced in his judging by receivers incorporating target material from previous trials into their responses. May believes this not to have been the case because (a) the receivers never commented on previous trials and (b) the judging task was never particularly difficult or subtle because the receivers were skilled at providing information that was both accurate concerning the target and unambiguous whether accurate or not and (c) the independent judge never made decisions about one trial based on information from another.

May points out that he informed R.W. about the nature of the judging process in August 1996, but that this information did not appear in our paper. However, our paper makes it quite clear that we were not suggesting that receivers explicitly mentioned previous targets on their response sheets or that the judge deliberately or knowingly made judging decisions across trials. Instead, we suggested a more subtle process wherein receivers might inadvertently include aspects of past targets in their responses without labelling them as such. The resulting response material would have been no more likely to be ambiguous in content and no more likely to involve subtle or difficult judgments than any other material. May believes that his judging decisions were not swayed by subtle cues but the large literature on implicit learning (see, e.g., Underwood, 1996) suggests that individuals can learn to detect and utilize very subtle cues without being aware that this is the case.

May does not raise the issue of the superior performance in static-target trials compared to that in dynamic-target trials to question the efficacy of this particular information leakage pathway but once more there are several ways in which such a difference could arise. For example, the individual images in the static targets might be more memorable and more likely to be reproduced by receivers in later trials; the similarities between responses and target material might be easier for the judge to detect in the single- image static targets; and so on.

Other Arguments against Sensory Leakage

Homogeneity of Effect Size

May points out that the effect size for Experiment One's static-target trials falls within the range of effect sizes (presumably of free-response ESP experiments) reported across a range of laboratories, experimenters, receivers and protocols. He draws no conclusion from this statement but its location in a section entitled "Other arguments against sensory leakage" appears to imply May considers it to be the basis of an argument against sensory leakage having been effective in Experiment One. It is difficult to surmise what that argument might be, however. The 78 published free-response studies meta-analysed by Milton (1997) that do not use altered states of consciousness for receivers - the nearest comparison group for Experiment One - report the use of, on average, just over half of the possible methodological safeguards that would be desirable, making their actual quality and any impact of quality on their outcomes hard to gauge. An effect size falling within the range of these studies' effect sizes is therefore no indication, let alone guarantee, that methodological problems left Experiment One unaffected.

Hypothesis confirmation and correlation

May argues that the proposed information leakage paths are implausible because a second experiment (Experiment Two) used a substantially different procedure and yet confirmed hypotheses developed in Experiment One. The two hypotheses involved simple effects: first, that performance on static targets in Experiment Two would be higher than static performance in Experiment One and second, that performance on dynamic targets in Experiment Two would be higher than performance on dynamic targets in Experiment One. Therefore, any artifact in Experiment Two that led to relatively high scoring overall would result in apparent confirmation of these two hypotheses. The relatively brief description of Experiment Two's procedure (Lantz et al., 1994), however, does not rule out a number of potential sensory information pathways. In addition, Experiment Two has not been subjected to a detailed critical analysis but our experience with assessing Experiment One calls into question whether the published details of Experiment Two are accurate and whether unrecorded procedures could be reliably reconstructed for analysis.

May also argues against information leakage paths having been effective in Experiment One because a statistically significant correlation between performance and the target's gradient of Shannon entropy (a measure of changes in luminance) was obtained in re-analyses of data from both Experiment One and Two (May et al, 1994). However, there is no reason to believe that these correlations count as evidence against the activity of methodological problems. First, there is no reason why artifacts cannot produce similar correlations between variables in different studies. In this case, targets with a high gradient of Shannon entropy may have visual characteristics that make them more striking or memorable or that identify themselves to the analyst as high-entropy targets, making them more likely to produce in both studies some of the types of leakage and judging biases outlined above. In addition, the evidence for the proposed relationship between the gradient of Shannon entropy and receiver accuracy appears weak. For example, in Experiment One the correlation appears to rely on the post hoc use of target ratings rather than ranks. In Experiment Two the correlation was only based upon data from three of the four receivers. May defends the exclusion of the fourth receiver by noting that the data from the other three receivers was individually significant. However, the post hoc correlation found in Experiment One was based upon the data of all receivers, regardless of individual significance and no reason is given for the change in data inclusion criterion.

Discussion

In this section we reply to each of the points mentioned by May in his discussion:

We did not ignore the difference in outcome between the static and dynamic conditions as May claims but clearly described it, together with supporting statistics, in our paper (Wiseman & Milton, 1998, p. 6). May's argument that limitation of statistically significant results to the static target conditions means that our proposed sensory leakage pathways were ineffective is incorrect, as discussed in the present paper.

We did not ignore information supplied by May regarding the judging procedures used in Experiment One, as he states. We did not report it because, as discussed above, this information did not affect the likelihood of the potential judging bias suggested in our paper. The leakage mechanism that May argues against is not the one that we proposed.

May criticises our ‘selective’ examination of Experiment One and our questioning of the other remote viewing studies from SAIC on the basis of this one assessment. However, as clearly stated at the start of our paper, Experiment One was examined because it was the only SAIC remote viewing study that had been both discussed (and endorsed) in detail in the AIR report and been published in a peer reviewed journal and therefore appeared likely to offer the most stringent methodological model for replication. The questions that arise from May's difficulties in providing a consistent account of its procedure apply to the other SAIC studies because the AIR assessors, like R.W., had, in part, to rely upon May to provide from memory extra details that had not been described in their written protocols (personal communication, Utts to R.W., 20 March 1996).

May criticises us for not describing the empirical research on the efficacy of sensory cues in ESP experiments, which, he claims, support his view that such cues would have been ineffective in Experiment One. The direct relevance of these studies to the conditions of Experiment One can be questioned but, contrary to May's claims, they demonstrate, as a group, that at least some cues can be exploited by even unselected participants under certain conditions.

May objects to our not noting the ‘replicated’ correlation between the gradient of Shannon entropy and receiver accuracy in Experiments One and Two. However, the evidence for such a replication appears open to question. May also criticises omission of a discussion of homogeneity of Experiment One's effect size with that of other, similar studies. As we have discussed above, having an effect size that falls within the range of a group of free-response

studies whose methodological quality is unclear is no guarantee that Experiment One was unaffected by methodological problems.

May criticises us for combining comparison groups and yet performed exactly the same analysis himself (Utts, 1995, p. 14) and reported it to Utts, who also reproduces it in her AIR report (Utts, 1995, p. 15). Whether it is, in fact, appropriate to combine data across any comparison groups depends upon the question that one is asking of the data. When one comparison group is shown to be independently statistically significantly above chance with a post-hoc analysis, as was the case with Experiment One's static target trials [Footnote 1], it is reasonable also to examine the combined results to get a general picture. However, our case does not rest on the fact that Experiment One's combined effect size is relatively low and we reported the static trials' outcome immediately after it to make it clear what effects we had to account for.

Personal commentary

May states that he agreed with R.W. to write a joint paper, along with Jessica Utts, concerning the potential for information leakage in Experiment One and was surprised to receive a draft of a paper by us, not listing himself or Utts as co-authors, a month before the deadline for the 1997 Parapsychological Association Convention. He states that he and Utts were requested to respond only with copy-style edits or to submit a separate paper if their responses were substantive. He states that because of time constraints they chose to do neither and that he spent the next month trying to establish what did, in fact, happen in Experiment One. He states that he had not previously given a high priority to obtaining accurate answers to R.W. 's questions on the Experiment One's procedure because he was expecting their publication to be joint.

R.W. did indeed have an agreement to write a joint paper with May and Utts and had agreed a format for the paper in which R.W. would describe any potential pathways for sensory information that he had identified, with May and Utts responding after each such description in the form of a dialogue (1 December 1996). Both May and Utts agreed to this format (personal

communication from Utts, 2 December 1997; personal communication from May, 4 December 1997). In a later email R.W. suggested that these comments could either be incorporated into the paper or could form a stand alone reply (14 February 1997). May acknowledged that he was happy with this (14 February 1997) and sent a copy of this correspondence to Utts, who raised no objection. Because J.M. had done considerable consultancy work on his project, R.W. credited her as a co-author of his descriptions of the methodological problems, and sent May and Utts a draft of the paper, consisting mainly of a list of potential information pathways as agreed, approximately one month before the deadline for the 1997 Parapsychological Association Annual Convention. May and Utts were not listed as co-authors on this draft because they had not yet indicated whether they intended to incorporate their replies into our text or have a stand-alone reply. R.W. repeated his invitation for May and Utts to comment either within the paper or separately (28 February 1997), offering to arrange for them to have a deadline extension beyond the remaining month if they needed it, in his capacity as program chair for the convention. May apparently saw R.W.'s offer as as an opportunity to only respond with copy-style edits or in a separate reply and did not take advantage of the offer of a deadline extension to prepare a written response.

R.W. made it clear from the beginning that his critique was intended for publication (20 March 1996) and May agreed to help him by supplying the information that he needed. However, May states that his anticipated co-authorship of a paper with R.W. made it unnecessary for him to give a high priority to making accurate the information that he gave R.W. about Experiment One. It is difficult to see what the point would be of R.W. producing a critique based on inaccurate information that May would then refute in part by producing a different version of events, if more careful attempts at reconstruction indicated such a version. Time spent by R.W. in attempting to assess Experiment One on incorrect information would be completely wasted.

Conclusion

We do not believe that any of the arguments raised by May rule out the potential problems in Experiment One as presented in our original paper, for the reasons we have given above. In addition, May does not address the two most crucial issues raised in our paper and we hope that he will make his position on these issues clear. They are as follows:

(1) We argue in our paper that the difficulties encountered by May and the SAIC team in producing a consistent account of several aspects of Experiment One's procedure make it difficult to be certain that the final collective account is reliable. As described in our paper, there are in total five different accounts of what happened and the differences between accounts are not trivial. Although May and his team clearly made considerable efforts to reach consensus about what actually happened, they appear to have done so after considerable discussion among themselves. Given that their memories appear to have been uncertain to begin with, it is possible that such discussion could lead the team to accept an inaccurate reconstruction simply because none of them could remember events clearly enough to contradict the final account. Memory does not function like a tape-recorder but can be influenced by information provided after the event. May, however, provides no documentation or argument as evidence to support his insistence that the final collective memory can be regarded as certainly reliable. If he has such evidence, we urge him to present it.

(2) As we point out in our paper, Utts often had to rely for her AIR report on May for details of undocumented aspects of the SAIC experiments. There are two reasons for believing that the information that May provided for her assessment may have been unreliable. First, the SAIC team's final collective account of Experiment One, that May insists is correct, is inconsistent in one identifiable respect with what he told Utts for her assessment, as documented in our paper. It cannot be the case that this final account is correct and that the information given to Utts is correct: one of them must be false and it follows that May must now be insisting that he misinformed Utts in this one respect. If he misinformed her in this one, identified respect in Experiment One, it is possible that he misinformed her about other unrecorded details of other

experiments in the SAIC programme. Second, the considerable problems experienced by May and the SAIC team in reconstructing a consistent account of a number of unrecorded details of Experiment One, suggests that similar problems may have arisen in other studies. As a result, it is far from clear whether Utts was in a position to make an accurate assessment of the SAIC work as a whole. Given the prominence in the public domain of Utts' generally favourable account of the SAIC work in the AIR report, this situation requires a response from May.

Footnotes

1) The ANOVA performed on Experiment One's data was performed on data combined across receivers. Lantz et al. (1994) note: "It is a general policy in the Cognitive Sciences Laboratory not to combine the data of receivers. In this study, all data combinations are post hoc" (p. 289).]

References

Palmer, J. (1983). Sensory contamination of free-response ESP targets: The greasy fingers hypothesis. *Journal of the American Society for Psychical Research*, 77, 101-113.

Palmer, J. and Kramer, W. (1986). Sensory identification of contaminated free-response ESP targets: Return of the greasy fingers. *Journal of the American Society for Psychical Research*, 80, 265-278.

Schmeidler, G.R. and Lindemann, C. (1966). ESP calls following an "ESP" test with sensory cues. *Journal of the American Society for Psychical Research*, 60, 357-362.

Underwood, G. (1996). *Implicit Cognition*. Oxford University Press: Oxford.

Wiseman, R. & Milton, J. Experiment One of the SAIC Remote Viewing Program: A critical reevaluation. *Journal of Parapsychology*, 62, 3-14, 1998.

Wiseman, R., Smith, M. D., & Kornbrot, D. Assessing possible sender-to-experimenter acoustic leakage in the PRL autoganzfeld. *Journal of Parapsychology*. 60, 97-128, 1996.

Department of Psychology
University of Hertfordshire
College Lane
Hatfield
AL10 9AB
England
psyqrw@herts.ac.uk

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh
EH8 9JZ
Scotland
UK 101335.2551@Compuserve.com