

**EXPERIMENTER EFFECTS AND
THE REMOTE DETECTION OF STARING**

Richard Wiseman

Perrott-Warrick Research Unit

University of Hertfordshire

UK

&

Marilyn Schlitz

Institute for Noetic Science

USA

**EXPERIMENTER EFFECTS AND
THE REMOTE DETECTION OF STARINGⁱ**

ABSTRACT

Each author recently attempted to replicate studies in which participants were asked to psychically detect an unseen gaze. RW's studies failed to find any significant effects whilst MS's study obtained positive findings. The authors then agreed to carry out the joint study described in this paper, in the hope of determining why they had originally obtained such different results. This joint study involved both MS and RW carrying out separate experiments, but running them in the same location, using the same equipment/procedures and drawing participants from the same subject pool. The studies involved placing experimenter and participant in separate rooms linked by a one way closed circuit television system. This allowed the experimenter to see the participant, but not vice versa. The experimental sessions were divided into two sets of randomly ordered trials. During 'stare' trials the experimenter directed his/her attention towards the participant; during 'non-stare' trials the experimenter directed this attention away from the participant. The participants' electrodermal activity (EDA) was continuously recorded throughout each session. Results revealed that the EDA of RW's participants was not significantly different during 'stare' and 'non-stare' trials. In contrast, the EDA

ⁱThe authors would like to thank the following organisations for supporting the research described in this paper; The Perrott-Warrick Fund, Institute for Noetic Sciences, UltraMind Ltd and the University of Hertfordshire. We are also grateful to Matthew Smith and Emma Greening for their help in running this experiment and analysing the data.

of MS's participants was significantly higher in 'stare' than 'non-stare' trials. The paper discusses the likelihood of different interpretations of this effect and urges other psi proponents and skeptics to run similar joint studies.

...the experimenter effect is the most important challenge facing modern parapsychology. It may be that we will not be able to make too much progress in other areas of the field until the puzzle of the experimenter effect is solved.

John Palmer, 1986, *Foundations of Parapsychology*, p. 220-221.

INTRODUCTION

The apparent detection of an unseen gaze (i.e., the feeling of being stared at, only to turn around and discover somebody looking directly at you) is a common type of ostensible paranormal experience, with between 68% and 94% of the population reporting having experienced the phenomena at least once (Coover, 1913; Braud, Shafer and Andrews, 1993a).

Some parapsychologists have attempted to assess whether this experience is based, at least in part, upon genuine psi ability. Such studies use two participants; a 'sender' and 'receiver'. These individuals are isolated from one another, but in such a way that the sender can see the receiver. Early experiments had the sender sitting behind the receiver (Titchener, 1898; Coover, 1913; Poortman, 1959), whilst later studies have employed one-way mirrors (Peterson, 1978) or closed-circuit television system (Williams, 1983; Braud, Shafer and Andrews, 1993a,b). The experimental session is divided into two sets of randomly ordered 'stare' and 'nonstare' trials. During 'stare' trials the sender directs his/her attention towards the receiver; during 'non-stare' trials the sender directs this

attention away from the receiver. Either during or after each trial a response is taken from the receiver. In early studies the receivers made verbal guesses as to whether they believed they had been stared at whilst later studies have measured receivers' electrodermal activity (EDA) throughout each trial. Many studies have obtained statistically significant differences between responses to 'stare' and 'non-stare' trials and a recent review of this work concluded that:

We hope that other investigators will attempt to replicate these studies. We recommend the design as one that is straightforward, has already yielded consistent positive results, and addresses a very familiar psi manifestation in a manner that is readily communicable and understandable to the experimental participants and to the public at large.(Braud, Shafer and Andrews, 1993b, p. 408).

Each of the authors recently attempted to replicate this 'staring' effect. The first author (RW) is a critic of parapsychology who wished to discover whether he could replicate the effect in his own laboratory. The second author (MS) is a psi proponent who has previously carried out many parapsychological studies and frequently obtained positive findings. The staring experiments carried out by RW showed no evidence of psychic functioning (Wiseman & Smith, 1994; Wiseman, Smith, Freedman, Wasserman & Hurst, 1995) whilst MS's study yielded significant results (Schlitz & LaBerge, 1994).

Such 'experimenter effects' are common within parapsychology and are open to several competing interpretations (see Palmer, 1989a,b). For example, MS's study may have

contained an experimental artifact absent from RW's procedure. Alternatively, MS may have worked with more psychically gifted participants than RW or been more skilled at eliciting participants' psi ability. It is also possible that MS and RW created desired results via their own psi abilities or fraud. Little previous research has attempted to evaluate these competing hypotheses. This is unfortunate, as it is clearly important to establish why experimenter effects occur, both in terms of assessing past psi research and attempting to replicate studies in the future. For these reasons the authors agreed to carry out a joint study in the hope of determining why our original studies had obtained such dramatically different results.

Our joint study involved MS and RW acting as separate experimenters for different sets of trials. The two sets of trials were carried out at the same time (early October 1995) and in the same location (RW's laboratory at the University of Hertfordshire in the UK). In addition, they used the same equipment, drew participants from the same subject pool and employed exactly the same methodological procedures. The only real difference between the trials was that one set was carried out by MS whilst the other set was run by RW. We were curious to discover if, under these conditions, we would continue to obtain significantly different results.

Method

Design

Both MS and RW carried out separate, but methodologically identical, studies concerned with the remote detection of staring. Each study had one independent variable which had

two levels ('stare' and 'non-stare'). Dependent variables consisted of participants' EDA during the experimental session and their responses to a 'belief in psi' questionnaire.

Participants

Thirty two subjects (10 males and 22 females; mean age of 25.72, age range 18-49) acted as receivers. Thirty participants were undergraduate psychology students studying at the University of Hertfordshire. The remaining two were the authors' colleagues. MS and RW acted as both experimenters and senders.

Apparatus and materials

Layout of rooms

It was clearly important to minimise the possibility of any sensory leakage between sender and receiver during the experimental sessions. For this reason the receiver was located in the University's Social Observation Laboratory whilst the sender was located in a small room approximately 20 meters away from the laboratory (see Figure 1).

Figure 1 about here

Video equipment

A Panasonic AG-450 video camera was positioned in front of the receiver and relayed an image (via a long cable connecting the two rooms) to a 14 inch JVC colour TV monitor

in the sender's room.

EDA measurement

Receivers' EDA was recorded using the RelaxPlus system (a commercially available hardware and software package produced by UltraMind Ltd). This system measures SRL (skin resistance level) by placing a constant current across two stainless steel electrodes and then recording the resistance encountered by that current at a rate of ten samples per second. The system filters for possible artifacts (caused, e.g., by movement) and records data to the computer's hard disk. The equipment (i.e., electrodes, input device, computer, computer monitor) was located next to the receiver throughout the experiment. The part of the programme involved in storing Ss details and physiological data could only be accessed via a password known only to MS and RW. Data from the RelaxPlus system was then fed into a spreadsheet (Microsoft's Excel) in order that the mean EDA for each 30 second trial could be calculated. All statistical analyses were carried out using the Statview software package.

Belief in psi questionnaire

Participants were asked three questions concerning their attitudes towards psi (see Appendix). Participants indicated their responses on a seven point scale ranging from -3 to +3. A general 'belief in psi' score was obtained by summing participant's responses over all three questions. Low scores on this questionnaire indicate strong belief in psi.

Trial randomization

Participants' EDA may decline during a session for several reasons (e.g., the apparatus measuring EDA warms up or participants may habituate to their surroundings). This decline could lead to artifactual evidence for psi if 'stare' trials tend to precede 'non-stare' trials. The following randomisation procedure was devised to minimise this possible artifact.

Prior to the experiment, an individual not involved in running the experiment (Matthew Smith) prepared a set of 32 sheets - each of which contained the order of the 32 'stare' or 'non-stare' trials for one session. For 16 of these sheets the trial orders were generated in the following way. MDS first opened the random number table (Appendix Three in Robson, 1983), chose a number as an 'entry point' into the table and then threw a die twice. The numbers obtained determined how he moved from this entry point to an actual 'starting point'. The eight consecutive numbers located in the row to the right of this starting point determined the order of the 'stare' and 'non-stare' trials. An even number translated into an ABBA (stare, non-stare, non-stare, stare) order whilst an odd number translated into a BAAB (non-stare, stare, stare, non-stare) order. The trial order for the remaining 16 sheets was determined by counter-balancing the orders of the randomised sheets just described. Thus, a 'stare, non-stare, non-stare, stare' on a 'randomised' sheet became a 'non-stare, stare, stare, non-stare' on a 'counter-balanced' sheet. All 32 sheets were then mixed together, placed into an opaque folder and kept in a locked drawer in RW's office. Matthew Smith was aware of the experimental hypotheses prior to carrying out the above randomisation procedure.

Procedure

Participants were run individually. On arriving at the laboratory, each participant was met by either RW or MS. Most participants were run by whichever experimenter was free to carry out the session, however, on a few occasions (e.g., when a participant was a friend or colleague of one of the experimenters) the experimenter would be designated in advance of the trial. Thus most participants were assigned to experimenters in an opportunistic, rather than properly 'randomised' (e.g., via random number tables or the output of an RNG), way. The experimenter showed them to the receiver's room and explained the purpose of the experiment. Next, the experimenter attached electrodes to the first and third finger of the participant's non-dominant hand and ensured that the RelaxPlus system was correctly monitoring their EDA. Participants were asked not to move their hand unnecessarily, nor to try to guess when they might be being stared at, but instead to simply remain as open as possible to any remote influence. The experimenter entered the participant's details into a computerised database, initiated the recording of EDA, started a stop-watch and left the receiver's room.

It was important that participants were not aware of the order of the 'stare' and 'non-stare' trials *before* the start of the experimental session. For this reason the list of trial orders was only selected by the experimenter *after* (s)he had left the receiver's room. The experimenter went to RW's office, retrieved the folder containing the lists of trial orders, selected any sheet they wanted and proceeded to the sender's room.

Two minutes after initiating the recording of the participant's EDA, the experimenter

started to carry out the designated order of 'stare' and 'non-stare' trials. The trial order was presented to the experimenters in the form of a list of 'stare' and 'non-stare' trials. During 'stare' trials the experimenter quietly directed his/her attention towards the receiver; during 'non-stare' trials the experimenter quietly directed this attention away from the receiver. Each trial lasted 30 seconds. Throughout this time the receiver completed the 'belief in psi' questionnaire and then read some magazines. All of the magazines were selected to be relatively bland in order to minimise possible effects on the receivers' EDA.

On completion of all 32 trials the experimenter returned to the receiver's room, thanked the participant and told them that they would be given feedback of the overall results within the next few weeks.

At the end of each experimental day, each experimenter copied that day's data (both from their own participants and the other experimenter's participants) onto their own floppy disk.

Resultsⁱⁱ

All analyses were preplanned.

ⁱⁱThis experiment was first reported at the 1996 Convention of the Parapsychological Association (Wiseman & Schlitz, 1996). Whilst preparing the paper for journal publication the authors reviewed the data and discovered an error in the way in which one participant's data had been transferred into the statistical package used for the analyses. For this reason the results reported here are slightly different to those reported in Wiseman

Primary analyses

A Wilcoxon signed rank test was used to compare participants' total EDA for the 16 'stare' trials with their total EDA during the 16 'non-stare' trialsⁱⁱⁱ. Participants run by RW did not differ from chance expectation (Wilcoxon $z=-.44$, $df=15$, $p[2\text{ tailed}]=.64$). In contrast, participants run by MS showed a significant effect (Wilcoxon $z=-2.02$, $df=15$, $p[2\text{ tailed}]=.04$).

A 'detect score' was then calculated for each participant by subtracting their total EDA during the 'stare' trials from their total EDA for the 'non-stare' trials. An unpaired t-test revealed that the detect scores of MS's participants were not significantly different from RW's participants ($df=30$, $t=1.39$, $p[2\text{ tailed}]=.17$).

Secondary analyses

Table 1 contains the correlation coefficients between participants' belief in psi questionnaire scores and their detect scores. Spearman rank correlation coefficients revealed that none of these correlations were significant. Table 1 also contains the means (and standard deviations) of the questionnaire scores for RW's group, MS's group and all

and Schlitz (1996).

ⁱⁱⁱPrevious studies (see, for example, Braud et al, 1993 a,b) have assessed their results by creating a 'psi score' (the sum of EDA during 'stare' trials divided by the sum of the total EDA) for each participant and then using a one sample t-test to determine the degree to which these scores deviate away from chance expectation. This procedure obscures whether an overall result is caused by a very small number of participants performing extremely well. The Wilcoxon sign rank test is more conservative than the one sample t-

participants.

| | RW's participants | MS's participants | All participants |
|--------------------|-------------------|-------------------|------------------|
| Mean | 1.94 | -.81 | .56 |
| Standard deviation | 4.22 | 4.12 | 4.33 |
| correlation (r) | -.15 | .32 | .15 |
| z value | -.58 | 1.23 | .84 |
| p-value | .56 | .22 | .39 |

Table 1 Means (and standard deviations) for the belief in psi questionnaire and the spearman rank correlation coefficients (corrected for ties) and p-values (2 tailed) between questionnaire scores and detect scores for RW's group, MS's group and all participants.

DISCUSSION

Participants run by RW did not respond differently to 'stare' and 'non-stare' trials. In contrast, participants run by MS were significantly more activated in 'stare', than 'non-stare' trials.

These findings can be interpreted in several ways.

First, one might argue that MS's significant results were caused by some type of

test as it is less influenced by the size of the deviation between participants' scores.

experimental artifact. Several steps were taken to guard against this possibility. For example, neither participants nor experimenters knew the order of the 'stare' and 'non-stare' trials before the start of the experiment, the location of the rooms minimised the possibility of any sender-to-receiver sensory leakage and the randomisation procedure ensured that the results were unlikely to be caused by progressive errors. This, coupled with the fact that one would expect any artifact to influence the results of both studies, suggests that MS's significant results are unlikely to have been caused by a methodological error.

Second, one could argue that either RW's or MS's results were caused by participant cheating. For example, participants could have discovered the order of 'stare' and 'non-stare' trials before the experimental session and altered their EDA accordingly. Alternatively, participants could have altered their data files so that they coincided with the order of 'stare' and 'non-stare' trials. Several factors mitigate against these ideas. First, such cheating would have been far from straightforward. For example, the selection of trial order was carried out a few moments before the start of the experimental session and it could only have been accessed by a participant who had installed some kind of covert monitoring equipment in the sender's room. Likewise, the computer could only be accessed if a participant had discovered a password which was known only to the experimenters. Also, neither RW's or MS's significant results are due to one exceptional participant and so one would have to hypothesize that several participants successfully cheated.

Third, the results could have been caused by experimenter fraud. Although the experiment was not designed to make such fraud impossible, its design does mean that certain types of cheating would have been extremely problematic. For example, neither experimenter could have decided to only include data from certain participants as the full list of all participants was known to both experimenters. However, more sophisticated forms of cheating were *theoretically* possible. For example, one experimenter could have substituted false sets of EDA values for participants' actual values before the data was analysed. Although possible, this would have been far from straightforward as participants were frequently scheduled back to back (thus cutting the time available for recording a false 'replacement' session to a minimum) and each experimenter made a back-up disk of all of the day's sessions at the end of each day (thus minimising the possibility of an experimenter substituting data after the day it had been recorded). In addition, no evidence of any cheating was uncovered during the running of the experiment or analysis of the data.

Fourth, one could argue that MS was working with a more 'psychically' gifted population than RW. This also seems unlikely as participants were assigned to the two experimenters in an opportunistic fashion.

Fifth, it is possible that MS was more skilled at eliciting participants' psi ability than RW. Interestingly, MS's participants scored higher on the 'belief in psi' questionnaire than RW's participants (although this difference just failed to reach significance - unpaired t value=1.86, df=30, p=.072 [2 tailed]). Given that participants were opportunistically

assigned to experimenters, this difference might be a reflection of the different ways in which RW and MS oriented participants at the start of the experiment. It seems quite possible that the experimenters' own level of belief/disbelief in the paranormal caused participants to express quite different levels of belief/disbelief in psi and have different expectations about the success of the forthcoming experimental session. Videotapes of RW's and MS's induction procedures are currently being analysed to identify differences in interaction and content.

Finally, it is also possible that both RW and MS used their own psi abilities to create the results he/she desired. This interpretation, if genuine, supports past research which suggests that 'successful experimenters' (i.e., those that consistently obtain significant effects in psi studies) outperform 'unsuccessful' ones on a variety of psi tasks (see Palmer, 1986 for a review of the literature supporting this notion).

In conclusion, this study reveals the value of developing collaborative relationships between skeptics and psi proponents. Both authors view this study as an initial step in the investigation of experimenter effects in psi research. Additional experiments would further aid our understanding of such effects. For example, it would be useful to carry out an experiment in which one experimenter interacted with the participant whilst the other carried out the 'stare' and 'non-stare' trials during the experimental session. Such a study would help discover whether our initial interactions with the participant, or behaviour during the experimental session, caused the results reported in this paper. The authors hope to carry out such a study in the near future, and urge other psi proponents

and skeptics to run similar studies.

REFERENCES

Braud, W., Shafer, D., & Andrews, S. (1993a). Reactions to an unseen gaze (remote attention): A review, with new data on autonomic staring detection, Journal of Parapsychology, **57**, 373-390.

Braud, W., Shafer, D., & Andrews, S. (1993b). Further studies of autonomic detection of remote staring: replications, new control procedures, and personality correlates. Journal of Parapsychology, **57**, 391-409.

Coover, J. E. (1913). The feeling of being stared at. American Journal of Psychology, **24**, 570-575.

Palmer, J. (1986). ESP research findings: the process approach. In H.L. Edge, R.L. Morris, J. Palmer, J.H. Rush (Eds.) Foundations of Parapsychology, Routledge & Kegan Paul: London. pp 184-222.

Palmer, J. (1989a). Confronting the Experimenter Effect. Parapsychology Review, **20** (4), 1-4.

Palmer, J. (1989b). Confronting the Experimenter Effect. Part 2. Parapsychology Review, **20** (5), 1-5.

Peterson, D. M. (1978). Through the looking glass: an investigation of extra-sensory

detection of being stared at. M.A. Thesis. University of Edinburgh.

Poortman, J. J. (1959). The feeling of being stared at. Journal of the Society for Psychological Research, **40**, 4-12.

Robson, C. (1983). Experiment, design and statistics in psychology. London, England: Penguin Books.

Schlitz, M. J. & LaBerge, S. (1994). Autonomic detection of remote observation: Two conceptual replications. In Bierman, D. J. (Ed.) Proceedings of Presented Papers, The Parapsychological Association 37th Annual Convention, The Netherlands, 352-360.

Titchener, E. B. (1898). The feeling of being stared at. Science, **8**, 895-897.

Williams, L. (1983). Minimal cue perception of the regard of others: the feeling of being stared at. Paper presented at the 10th Annual Conference of the Southeastern Regional Parapsychological Association, West Georgia College, Carrollton, GA. See Journal of Parapsychology, **47**, 59-60.

Wiseman, R. & Smith, M. D. (1994). A further look at the detection of unseen gaze. Proceedings of the Parapsychological Association 37h Annual Convention, Parapsychological Association. 465-478.

Wiseman, R., Smith, M.D., Freedman, D., Wasserman, T., & Hurst, C. (1995). Two further experiments concerning the remote detection of an unseen gaze. Proceedings of the Parapsychological Association 38h Annual Convention, Parapsychological Association. 480-490.

Wiseman, R. & Schlitz, M. (1996). Experimenter effects and the remote detection of staring. Proceedings of the Parapsychological Association 39h Annual Convention, Parapsychological Association. 149-157.

APPENDIX

Belief in psi questionnaire

Please use the following definition for the three questions that follow.

Psi: Direct interactions between mental processes and the physical world or other mental processes occurring outside currently understood channels. Thus this is a 'blanket' term used to refer to all paranormal processes and causation.

1. Is the existence of psi :

Certain -3 -2 -1 0 +1 +2 +3 *Impossible*

2. What best describes your own psi ability?

I have psi ability -3 -2 -1 0 +1 +2 +3 *I have no psi
ability*

3. Do you believe you might be able to demonstrate any psi ability in this experiment?

Yes -3 -2 -1 0 +1 +2 +3 *No*

