University of Hertfordshire

Experimenter Effects in Parapsychology: Replication and Mechanism[1]

Richard Wiseman

Department of Psychology, University of Hertfordshire

Caroline Watt

Department of Psychology, University of Edinburgh

Experimenter Effects in Parapsychology: Replication and Mechanism[1]

Richard Wiseman

Department of Psychology, University of Hertfordshire

Caroline Watt

Department of Psychology, University of Edinburgh

[1] Acknowledgements. We are grateful to the Fundaçao Bial for funding this study. Thanks to Emma Greening and Ciarán O'Keefe for assisting with participants in the study, and to Helen Large for rating the video-tapes. Thanks to Professor Harvey Irwin and Professor Ian Deary for their helpful comments on an earlier draft.

....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, *Foundations of Parapsychology*, (Edge, Morris, Palmer & Rush, 1986), p.220-221.

Introduction

Rosenthal (1976) has carried out extensive research into the possible influence of experimenter belief and expectation on study outcome. In his early research, Rosenthal found that psychologists with differing expectations about their participants tended to obtain results in line with their expectations. The results clearly demonstrated that the experimenters' expectations were increasing the probability of their participants behaving as expected (Rosenthal & Rubin, 1978). Subsequent research revealed that such "experimenter effects" were not confined to the behavioural sciences, but could also affect the outcome of physical and biological research (see Wiseman & Watt, 1999).

Recent work on the experimenter effect has moved away from simply demonstrating the existence of the phenomenon, and towards examining its practical applications and theoretical implications (Rosenthal, 1990). For example, studies examining the relationship between experimenter belief and study outcome have helped explain why certain areas of scientific research have yielded apparently inconsistent results (Luborsky et al., 1999). Other work has started to explore some of the possible underlying mechanisms for the effect by examining the types of procedures that are most susceptible to the influence of experimenter belief (Harris & Rosenthal, 1985; Rosenthal & Rubin, 1978). There has also been research on possible personality moderators of experimenter effects (Cooper & Hazelrigg, 1988; Hazelrigg, Cooper & Strathman, 1991).

The present study builds upon this work by exploring the possible existence of experimenter effects within two important areas of parapsychology - the psychology of belief in the paranormal and psi research. The study explores the degree to which seemingly inconsistent results within each of these areas can be explained by experimenter effects, and examines the types of participant-experimenter interactions associated with the effect. The experiment is made up of two strands and the background to each aspect of the study will be described in turn.

Cognitive correlates of belief in the paranormal.

Some psychologists have examined whether belief in the paranormal correlates with performance on various cognitive tasks including critical thinking, reasoning skills, and IQ (see reviews by Irwin, 1991, 1999; French, 1992). Although findings are mixed, the overall trend is that people who disbelieve in the paranormal tend to outperform believers on these tasks (Blackmore & Troscianko, 1985; Wiezerbiki, 1985; Roberts & Seager, 1999), and Irwin has named this the "cognitive deficits" hypothesis. However, many of these studies have been conducted by experimenters who are skeptical about the existence of the paranormal (e.g. Alcock & Otis, 1980), and some experimenters who are more open to the notion of psi have failed to replicate these results (e.g., Irwin, 1991; Roe, 1999). This has led some researchers (e.g. Irwin, 1991; Smith, Foster & Stoven, 1998) to hypothesise that these inconsistent results may be due, at least in part, to the experimental context influencing participants' reports concerning their level of belief in the paranormal, and/or their performance on cognitive tasks.

Research investigating how the belief-cognitive ability relationship may be influenced by experimental context has had mixed results. Merla-Ramos (2000) found that paranormal believers tended to have poorer syllogistic reasoning ability than disbelievers, but only for syllogisms that included a paranormal or religious content. No belief-reasoning correlation was found for syllogisms that had a neutral content.

Irwin (1991) used a similar syllogistic reasoning task, including neutral, pro-paranormal and anti-paranormal syllogisms. Overall he found a non-significant syllogisms-belief correlation, and the content of the syllogisms did not affect the correlation. Smith, Foster & Stovin (1998) varied the context in which participants completed a paranormal belief questionnaire and did an intelligence test (Ravens Advanced Progressive Matrices), with some participants being given a psi-supportive context and some a psi-unsupportive context. This study found that participants expressed greater paranormal belief in the psi-supportive condition. Overall a negative correlation was found between belief and performance on the matrices task. However, contrary to prediction, context did not appear to affect this correlation.

The present study investigates the effect of experimental context on participants' paranormal belief and cognitive ability by having two experimenters with differing attitudes towards the paranormal (CW and RW) carry out an experiment examining whether belief in the paranormal correlates with performance on two cognitive tasks. CW has conducted many empirical investigations of possible paranormal cognition (e.g., Watt, 1996; Watt, Ravenscroft, & McDermott, 1999) and is somewhat less skeptical about the paranormal than RW, who is a well-known critic of parapsychology (e.g., Wiseman, 1997). In order to build upon the previous literature reviewed above, the two tasks selected to assess cognitive ability are a syllogistic reasoning task (as used by Irwin, 1991) and Ravens Matrices (as used by Smith, Foster & Stovin, 1998). To facilitate comparisons with these earlier studies we also use the same belief questionnaire as they did, and we use correlational analyses as they did. We also film the interaction between experimenters and participants in order to provide additional information about differences between the experimenters, for example in the extent to which they conveyed belief or skepticism about psi.

Two possible mechanisms may operate to produce a correlation between belief in the paranormal and cognitive ability. Firstly, it is possible that participants are shifting their responses on the belief questionnaire, perhaps in line with demand characteristics

from their experimenter, with the more intelligent participants being more perceptive and responsive to these subtle experimenter cues. Secondly, it is possible that participants are altering their performance on the cognitive tasks, for instance perhaps under certain circumstances believing participants are less motivated to make an effort to score well on the cognitive task. The present study will enable the exploration of such possible mechanisms underlying the belief-cognitive ability correlation.

The experimenter effect in psi research.

Certain parapsychologists have become seen as "psi-facilitators" because they consistently carry out studies that obtain evidence for psi, whilst others have become known as "psi-inhibitors" because they consistently obtain chance results (Schmeidler, 1997). Researchers have outlined many different explanations that may account for this effect (see Palmer, 1989a, b). Some have suggested that psi-facilitators may be carrying out their experiments with psychic participants and/or be employing experimental procedures that enhance participants' psychic ability. Others have suggested various "normal" explanations for the effect including, for example, that psi-facilitators are fraudulent or that their experimental protocols are flawed.

To help resolve the issue, Dr Richard Wiseman (RW) recently conducted a joint study with parapsychologist Dr Marilyn Schlitz (MS) from The Institute of Noetic Sciences in California. MS holds a strong belief in the existence of psi and has consistently obtained positive psi results. In contrast, RW is skeptical about psi and has failed to obtain evidence of psychic functioning. In 1995 MS and RW carried out a joint study to help determine why they had obtained such different results. The studies were conducted at the same location, and used the same participant pool, equipment and procedure. In each study participants were asked to psychically influence the physiological activity of another person located in a distant location. Half of the trials were conducted by MS whilst the other half were carried out by RW. Results revealed evidence of an experimenter effect, with MS's trials showing a significant psi

effect, whilst RW's trials were at chance (Wiseman & Schlitz, 1998). This experiment was later repeated and obtained the same pattern of results (Wiseman & Schlitz, 1999).

These findings suggest that experimenter effects in psi research may not be due to differences in participant population, experimental procedures, or fraud, but support the notion that the experimenter's beliefs and expectations may indeed influence study outcome. The second strand of the present study explores whether this effect can be replicated in another joint experiment carried out by a different experimental team (CW and RW) and employing a remote viewing ESP task. Based on the pattern found in Wiseman and Schlitz's studies, we predict that RW's participants will have lower psi scores than CW's participants. Filming of the interaction between the experimenter and the participant during the ESP task will enable exploration of how various factors may correlate with psi target rank.

Both types of effect have important implications. If CW and RW find their participants differ in either belief in the paranormal and/or performance on cognitive tasks, this would have important implications for mainstream cognitive psychology, and would question the degree to which well established effects relating to belief and cognition may be due to experimenter effects. It might also throw light on the presently inconsistent findings pertaining to the cognitive deficits hypothesis. Secondly, if experimenter effects were found for the psi task, it would strongly suggest that RW's past findings were not due to chance, or confined to one parapsychologist or a single type of psi task. Instead, it would suggest that experimenter expectation was a significant, and more general, factor in determining the outcome of psi research. Such a finding would help explain why certain experimenters have obtained consistent evidence for psi whilst others have not. More importantly, the finding would also greatly assist those attempting to develop experimental procedures that produce replicable evidence for psi.

Method

Participants.

The volunteer participants were mostly undergraduate students at the University of Hertfordshire, who received either £5 or course credit for taking part. 30 participants were tested by RW and 30 were tested by CW. There were 19 male and 40 female participants (sex data were not available for one participant); mean age was 24.9 (SD = 8.2) (age data were not available for three participants).

Materials.

Belief in the Paranormal Questionnaire.

This was Tobacyk's Revised Paranormal Belief Scale (Tobacyk, 1988; Lange, Irwin & Houran, 2000), a 26-item questionnaire containing items measuring a variety of paranormal beliefs. Participants respond to each item on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Scores could range from a minimum of 26 to a maximum of 182. The overall scale consists of 7 sub-scales: traditional religious beliefs, psi, witchcraft, superstition, spiritualism, extraordinary life forms and precognition.

Ravens Matrices Test.

Set 1 of Ravens Advanced Progressive Matrices was used to assess non-verbal reasoning ability (Raven, Court & Raven, 1985). This was a 12-item task. For each item, participants were required to indicate which of eight possible symbols correctly completed a sequence of symbols. Participants were given 5 minutes to complete this task. Scores could range from a minimum of 0 to a maximum of 12.

Syllogisms Test.

This test of reasoning skills contained 24 items (see Appendix 1), and was an abbreviated version of the 48-item test developed by Irwin (1991), omitting the 24 parapsychological items. Each item contained a pair of statements, followed by a

conclusion. The participant's task was to indicate whether the conclusion was valid or invalid, as derived logically from the statements. Participants were given 5 minutes to complete this task. Scores could range from a miminum of 0 to a maximum of 24.

Psi Targets.

A single target pool of five video-clips, each approximately two-minutes in duration, was used for the psi task. Each clip was on a separate video cassette. Each clip had a different instrumental musical soundtrack. These targets were prepared in line with suggestions from previous literature that free-response targets should be dynamic, individually thematically coherent, maximally different from one another, and have a neutral emotional tone without being bland (Delanoy, 1988; Watt, 1988). The same set of targets had been used in a previous remote viewing study (Wiseman & Greening, 1998).

Testing room.

The experiment took place in a quiet, comfortable and spacious laboratory in the Department of Psychology at the University of Hertfordshire. The room contained sofa-style seating, two tables and chairs, chairs for the experimenter, and a comfortable reclining chair. There was also a video player and monitor, facing the reclining chair. The room contained three unobtrusive remotely-controlled cameras, two of which were used to film the experimenter and the participant during the session. This filming was monitored and controlled by a technician in an adjacent sound-attenuated control room. There were windows on three walls of the room, all of which were covered by blinds. The lighting of the room could be varied to give brighter lighting for completing questionnaires and dimmer lighting for the psi task. While waiting to do the psi task, participants were taken to a separate suite of rooms located about 60 feet away, so that they were sensorially isolated from the main testing room.

Procedure

Target randomisation procedure.

Prior to the study beginning, a person otherwise unconnected with the study used published random number tables to designate a target clip for each participant (Rand Corporation, 1955). A calculator RND function was used three times to select: 1. A page of random numbers (one of ten); 2. A column of numbers; 3. A row, to give an entry point to the table. The randomiser read from the selected entry point until a digit from 1 to 5 was encountered. This digit was then transformed to the letters A to E according to a rotating code, such that even if a single digit unexpectedly occurred more frequently than others throughout the study, this would translate to different target identities. The selected target clip (labelled A, B, C, D or E) was written on a slip of paper, that was sealed in a manilla envelope which was then folded and sealed inside a white envelope labelled with the participant number. The record of targets selected was sealed by the randomiser in a plastic tamper-evident security bag as have been used in previous ESP research with take-home targets (Delanoy, Watt, Morris & Wiseman, 1993), and the randomiser had no contact with the experimenters or anyone else involved in the experiment after the randomisation had taken place. The random numbers used to produce each target were also recorded by the randomiser so that the steps to selecting the target could be recreated after the study, if necessary. Also, the record made by the experimenter of each session's target could be double-checked against the randomiser's record to verify that the correct target had been played.

The study took place over five days and participants were allocated to experimenters in a counterbalanced fashion. Participants (a maximum of three per session) were greeted on their arrival by an assistant who brought the participants to the testing room to meet the experimenter. The participants were seated on the sofas while the experimenter briefly explained the procedure to them. The experimenter also informed the participants that parts of the session would be filmed unless anyone objected

(there were no objections to the filming). During this initial chat, the experimenter was filmed.

In order to study experimenter differences, it was decided to adopt a naturalistic method, whereby RW and CW conducted the session in whatever way each felt most natural and comfortable. There was no explicit attempt to communicate the experimenter's psi-openness or skepticism to participants, as neither experimenter felt they would naturally do this in a study. However, CW stated that her main area of interest was psi research and that this was her career, and that she found this research interesting and enjoyable. The implicit message was intended to be that she valued this area of research and felt positively towards it. RW's initial chat was very brief and said nothing explicit about his beliefs about psi or his reasons for doing the study. However, it was clear from many participants' responses when they met RW that they recognised him from his appearances in the media, and thus knew his reputation as a critic of the paranormal.

Each set of questionnaire materials was numbered with a participant identification number and put on a clipboard prior to the start of the session. These clipboards were distributed randomly to participants following the initial chat. To complete the questionnaires, two participants were then seated at the tables while the third remained seated on the sofa, in order that participants' responses remained private. Participants first completed the belief questionnaire, as directed by the experimenter. They then completed either the Syllogisms task or the Matrices task, in a counter-balanced fashion (i.e. in session 1, syllogisms was completed first, in session 2, matrices was completed first). Following completion of the questionnaire measures, the participant with the lowest participant identification number ("participant 1") remained with the experimenter in the testing room while the assistant took the remaining two participants ("participants 2 and 3") along to the waiting area.

The experimenter dimmed the lights and seated participant 1 in the comfortable chair for the psi task. Filming of the experimenter and the participant commenced at this point, though this was not drawn to the participant's attention. The experimenter sat beside the participant with a clipboard, and explained the psi task.

The psi task.

The psi task in the present study was a free-response ESP task using an "interview technique" similar to that used in remote viewing research (e.g. Targ & Puthoff, 1977). The psi task was novel in that the experimenter judged the participant's impressions and produced target rankings during the session, immediately following the interview with the participant. Instant judging was possible because prior to the study the experimenters had familiarised themselves with each of the five possible targets.

The participant was seated in a comfortable chair facing a video screen, and the experimenter, who was blind to the target identity, sat adjacent to the participant. The experimenter presented the task as a precognitive one, that is, that the participant would in a few minutes be viewing the randomly-selected target clip. The participant was asked to give his or her impressions of the clip that would shortly be viewed. This was done without any prior relaxation or altered state induction. Instead, as with remote viewing methods, the target mentation was recorded in the form of a brief interview between the experimenter and the participant. The experimenter asked questions to elicit additional information from the participant and wrote down the participant's responses.

After about five minutes of interview, the experimenter judged the mentation items against the five target possibilities and ranked the targets in order of similarity to the mentations. The experimenter's ranking choice was written on the session record sheet and was stated aloud so that it was recorded on video. Then the experimenter retrieved the envelope containing the target identity (prior to the start of the session, RW had placed this envelope out of sight on top of the video monitor, CW had placed

it in her pocket). For CW's sessions, the participant was invited to open the envelope and show the slip containing the target identity to CW. For RW's sessions, RW opened the envelope containing the target identity. Video recording ended once the experimenter had revealed the target identity and noted it on the session record sheet. The experimenter then selected and played the appropriate video cassette containing the target clip. The psi task took about 10 minutes including viewing of the target clip. The participant never saw any of the other target possibilities. Although the task was presented to the participant as precognitive, this was not a true precognition task as the target identity had been selected prior to the participant generating his or her mentation.

The experimenter then debriefed the participant about how the psi task would be scored and how the questionnaire measures would be correlated with the psi task. Participant 1 was asked not to discuss the target clip with other potential participants and was then free to leave.

Meanwhile, participant 3 waited for another 10 minutes while the experimenter went with participant 2 to the testing room. The psi task took place as already described. Participant 2 was then free to leave, and the experimenter brought participant 3 along to the testing room for the psi task. The session ended after participant 3 had completed the psi task.

The distribution of participant groups was equal for each experimenter: each did 7 sessions with 3 participants, 4 sessions with 2 participants, and one session with one participant. The session length when there were three participants was approximately one hour (approximately 30 minutes for questionnaires, approximately 30 minutes for psi testing). There were no drop-outs from the study.

Results

The experiment terminated when, as planned, each experimenter had tested 30 participants.

Missing data.

One participant did not complete the syllogisms task, due to language difficulties, therefore N = 59 for this questionnaire. One participant misunderstood the instructions for doing the matrices task, so N = 59 for this measure. On double-checking whether the correct target had been played during each session, it was found that in two sessions the wrong target had been played because CW had misread the hand-written target identification letter. Before doing any analysis, we decided to include the data from these two sessions "at face value", since the participant's task was to identify the target to be shown at the end of the session. However, we will also present (Appendix 2) the psi data analysed excluding these two sessions, for the sake of completeness. As can be seen from Appendix 2, this ambiguity has no material effect on the results.

Manipulation check.

The experimenters' initial conversation with participants was filmed, so it was possible to independently rate these recordings for the degree of skepticism or psi belief that the experimenter portrayed. This would enable us to verify whether RW did indeed appear to be more skeptical about psi than CW, as had been intended. To do the rating, an individual who was blind to participants' scores on the belief, cognitive and psi measures watched each video-recording of the initial experimenter chat with participants and rated the experimenter's apparent psi belief/skepticism on a seven point scale, ranging from 1 (proponent) through 4 (neutral) to 7 (skeptic). A group-level analysis was used here because participants were tested in 24 groups ranging from one to three individuals. RW's mean skepticism rating was 4.7 (SD = .78), CW's mean skepticism rating was 4 (SD = .43). Using a Mann-Whitney test, RW was rated as significantly more skeptical than CW ($Z_{\text{corrected for ties}} = 2.35$, $p = .02$,

two-tailed), confirming the validity of the experimental manipulation. However, CW was rated as neutral in belief for most sessions and this may suggest that her sessions would not be directly comparable to those for a strong psi-proponent, such as Marilyn Schlitz.

Belief in the paranormal questionnaire.

Table 1 presents mean scores for the belief in the paranormal questionnaire, overall and broken down by experimenter. The table includes t-tests of the difference between belief scores for CW's versus RW's participants, and shows that there is no difference between the experimenters' belief scores, either overall or for any of the belief sub-scales. Therefore there is no indication that participants have shifted their responses on the belief questionnaire in line with the experimenters' own attitudes towards psi.

BELIEF	Overall	CW	RW	t
Overall belief score	97.7 (24.1)	97.6 (24.3)	97.9 (24.4)	.04 (.97)
Traditional religious beliefs	19.7 (6.8)	20.3 (7.2)	19.1 (6.5)	-.68 (.50)
Psi	15.7 (5.4)	15.1 (5.5)	16.3 (5.4)	.83 (.41)
Witchcraft	15.0 (5.8)	15.0 (5.2)	15.1 (6.5)	.11 (.91)
Superstition	6.9 (4.7)	7.3 (4.8)	6.5 (4.6)	-.69 (.49)
Spiritualism	15.5 (5.0)	15.6 (5.0)	15.4 (5.2)	-.13 (.90)
Extraordinary life forms	10.4 (3.3)	9.8 (3.0)	11.0 (3.5)	1.46 (.15)
Precognition	14.5 (5.1)	14.5 (5.1)	14.4 (5.2)	-.08 (.94)

Table 1. Mean and standard deviation (in parentheses) of scores on the Revised Paranormal Belief Scale, overall and by experimenter. The final column gives t-tests of the difference in mean belief scores for CW vs RW, with two-tailed p-values in parentheses.

Task	Overall	CW	RW	t
Ravens Matrices	8.9 (1.8)	8.6 (2.1)	9.1 (1.3)	.90 (.37)
Syllogisms	12.0 (3.5)	11.4 (3.8)	12.8 (3.0)	1.54 (.13)

Table 2. Mean and standard deviation (in parentheses) of scores on the cognitive tasks, overall and by experimenter. The final column gives t-tests of the difference in mean cognitive scores for CW vs RW, with two-tailed p-values in parentheses.

Cognitive tasks.

Descriptive data for the Ravens Advanced Progressive Matrices and Syllogisms tasks are presented in Table 2, overall and by experimenter. The table includes t-tests of the difference in cognitive scores for CW's versus RW's participants. Though RW's participants tend to have higher scores on the cognitive tasks than CW's, this difference is not statistically significant.

Experimenter effects for cognitive tasks and belief in the paranormal.

In order to explore the question of whether the two experimenters would obtain different patterns of correlation between participants' psi belief and performance on the cognitive tasks, Pearson correlation coefficients were calculated and are presented in Table 3 (Ravens Matrices) and Table 4 (Syllogisms). Overall, there is no correlation between performance on the Matrices and belief in the paranormal, neither is there any correlation for the experimenters individually. This does not confirm the previous finding by Smith, Foster & Stovin (1998) of a negative correlation between belief and Matrices scores.

BELIEF	Overall	CW	RW
Overall belief score	-.04 (.74)	-.06 (.74)	-.02 (.94)
Traditional religious beliefs	-.09 (.48)	-.17 (.39)	-.06 (.73)
Psi	-.05 (.72)	-.11 (.58)	.01 (.95)
Witchcraft	.08 (.57)	.03 (.88)	.14 (.46)
Superstition	-.10 (.46)	.04 (.83)	-.31 (.09)
Spiritualism	.09 (.48)	-.06 (.74)	-.14 (.45)
Extraordinary life forms	.19 (.16)	.18 (.35)	.18 (.36)
Precognition	-.06 (.66)	-.07 (.70)	-.04 (.83)

Table 3. Pearson correlation coefficients (with two-tailed p-values in parentheses) for belief in the paranormal and scores on Ravens Matrices, overall and by experimenter.

As Table 4 shows, overall, there is a statistically significant negative correlation between belief in the paranormal and performance on the syllogisms task ($r = -.28$, $p = .03$, 2-t). That is, those with greater belief in the paranormal tend to have lower scores on the syllogisms task. The breakdown of correlations by experimenter reveals that this correlation is due to CW's participants. RW's participants score close to chance, while CW's have an independently significant negative correlation between belief and syllogisms scores ($r = -.45$, $p = .01$, 2-t). Looking at the belief sub-scales, none of the belief-syllogism correlations for RW's participants is statistically significant. For CW's participants, the traditional religious beliefs, spiritualism, and precognition sub-scales are all independently significant. This seems to show clear evidence of an experimenter effect, in that CW's participants generally show a negative belief-syllogisms correlation, while RW's show no belief-syllogism correlation. The difference between the experimenters' correlations (calculated using Fisher Z transformation) is statistically significant for traditional religious beliefs ($p = .03$, 2-t) and spiritualism ($p = .01$, 2-t).

BELIEF	Overall	CW	RW
Overall belief score	-.28 (.03)*	-.45 (.01)*	-.08 (.70)
Traditional religious beliefs	-.32 (.01)*	-.53 (.00)*	.02 (.92)
Psi	.01 (.91)	-.03 (.87)	.03 (.89)
Witchcraft	-.25 (.06)	-.28 (.14)	-.26 (.18)
Superstition	-.24 (.07)	-.18 (.35)	-.30 (.12)
Spiritualism	-.19 (.14)	-.48 (.01)*	.16 (.42)
Extraordinary life forms	.17 (.20)	.00 (.98)	.28 (.13)
Precognition	-.32 (.01)*	-.45 (.01)*	-.16 (.39)

Table 4. Pearson correlation coefficients (with two-tailed p-values in parentheses) for belief in the paranormal and scores on the Syllogisms task, overall and by experimenter. Correlations that are significant at or below the .05 level are indicated with an asterisk.

Relating our overall belief-syllogisms correlations to previous research using the same measures, it is interesting to note that our pattern of belief-syllogisms correlations replicates that found by Irwin (1991). The two strongest negative belief-syllogisms correlations found by Irwin were for the traditional religious belief and the precognition sub-scales (with a statistically significant correlation for the former). The only two independently significant correlations for our sub-scales were for traditional religious belief and precognition. Our data therefore confirm Irwin's findings on these two sub-scales. For the full PBS scale, our findings are also in line with Irwin's, as he also found a negative correlation between belief and syllogisms scoring, though Irwin's correlation was not statistically significant.

Psi task.

Table 5 shows the distribution of target ranks for all 60 sessions. The planned analysis for the psi task was sum-of-ranks. This was calculated for the data in Table 5 using the formula in Solfvin, Kelly & Burdick (1978, p. 99). The sum-of-ranks was 173 (compared to MCE sum of ranks of 180), which gives an almost exactly at chance Z = .05 (adopting the convention of giving a positive sign to the Z when the data is in the psi-hitting direction). Therefore there is no evidence that the target video clip was identified more often than chance.

	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**
Actual	17	10	10	9	14
MCE	12	12	12	12	12

Table 5. Rank allotted to target.

	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**
CW	6	6	5	6	7
RW	11	4	5	3	7

Table 6. Rank allotted to target, for CW and RW separately.

Table 6 shows the distribution of target ranks separately for the two experimenters. The mean target rank for CW was 3.1 (SD = 1.4), compared to a mean of 2.7 (SD = 1.6) for RW. This difference was not statistically significant on a Mann-Whitney test ($Z_{\text{corrected for ties}} = -1.067$, p = .286, two-tailed). Therefore there is no indication of an experimenter effect on psi results. This does not replicate the previous findings of Wiseman and Schlitz of an experimenter effect on psi, indeed the trend is in the opposite direction, with RW obtaining slightly better psi scores than CW.

Belief in the paranormal and psi.

Spearman correlation coefficients were carried out between overall belief in the paranormal and psi target rank, and between each of the 7 belief sub-scales and psi target rank. Non-parametric statistics were used in this case because it was felt the target rank data did not meet the assumptions for parametric testing. The results are shown in Table 7, both overall and broken down by experimenter. Overall, there is no sign of a relationship between the participants' paranormal belief scores and performance on the psi task. Also, the two experimenters do not differ in terms of their participants' belief-psi correlation.

BELIEF	Overall	CW	RW
Overall belief score	.06 (.67)	.04 (.82)	.07 (.72)
Traditional religious beliefs	.09 (.50)	.07 (.70)	.08 (.65)
Psi	.01 (.92)	-.22 (.23)	.26 (.16)
Witchcraft	.06 (.63)	.18 (.33)	-.01 (.94)
Superstition	.00 (1.0)	.16 (.40)	-.14 (.46)
Spiritualism	.03 (.82)	-.18 (.31)	.23 (.22)
Extraordinary life forms	.05 (.70)	.03 (.87)	.16 (.40)
Precognition	-.01 (.96)	.06 (.75)	-.06 (.72)

Table 7. Spearman correlation (corrected for ties, two-tailed p-values in parentheses) between psi target rank and belief in the paranormal, overall and broken down by experimenter.

Independent rating of experimenter-participant interactions.

The psi task in the present study involved an interview technique where the experimenters attempted to elicit target-related information from participants. These interviews were filmed and therefore could be rated to explore the relationship between characteristics of the interviews and psi target rank.

An independent rater, who was blind to the outcome of the psi task, rated each of the 60 experimenter-participant interviews on three variables that might relate to psi-success: degree of spontaneity of the participants' utterances (rated on a 7 point scale from 1 = not at all spontaneous to 7 = very spontaneous); quantity of images/ideas that the participant expressed (rated on a 7 point scale from 1 = small quantity of images/ideas to 7 = large quantity of images/ideas); and amount of detail in the participants' statements (rated on a 7 point scale from 1 = small amount of detail to 7 = large amount of detail). Spearman correlations were then calculated between the ratings and the target ranks. All correlations were near-zero (spontaneity $r_s = .026$, $p = .84$, two-tailed; quantity $r_s = -.086$, $p = .51$, two-tailed, detail $r_s = -.034$, $p = .79$, two-tailed) therefore there was no indication that psi target rank was related to any of these factors.

Discussion

This study explored whether experimenter effects might influence a correlation between belief in the paranormal and performance at cognitive tasks, by comparing said correlations for RW's participants and CW's participants. Two cognitive tasks were used: Ravens Matrices, and a syllogisms task. Neither experimenter found any correlation between performance on the Matrices task and belief in the paranormal, which is inconsistent with the findings of a previous similar study (Smith, Foster & Stovin, 1998). However, there did seem to be evidence of an experimenter effect for the syllogisms task, with RW's participants showing no belief-syllogisms correlation, while CW's participants had a significantly negative belief-syllogisms correlation. That is, for CW those with greater belief in the paranormal scored relatively poorly on the syllogisms task.

How are we to interpret the observed pattern of a negative correlation between belief and syllogisms performance for CW but not for RW? Irwin (1991) argues that the

previous research showing a negative correlation between belief and cognitive skills is mostly for skeptical experimenters. He suggests that perhaps cognitively-skilled participants are inhibiting their admission of paranormal belief when tested by skeptical experimenters. Our results do not support this interpretation because we found no difference between the two experimenters in obtained paranormal belief scores, despite the fact that RW was independently rated as more skeptical in his interactions with participants than CW. However, it is possible that our participants are shifting their syllogisms scores in response to their experimenter. RW's participants had a mean syllogisms score of 12.8, which is non-significantly higher than the mean of 11.4 for CW's participants (unpaired $t[57] = 1.545$, $p = .13$, 2-t). This may suggest that some participants have slightly shifted their syllogisms performance, perhaps due to increased or decreased motivation to do well. Therefore in terms of overall mechanisms underlying the belief-cognitive performance correlation, our data may indicate that participants are varying their performance on the syllogisms task, rather than varying their responses on the belief questionnaire.

In order to investigate this question further, all 60 participants were divided into believers and disbelievers on a median split. T-tests verified that there was no indication that participants had shifted their belief scores for the different experimenters because there was no significant difference between the belief scores of RW's believers and CW's believers (RW mean belief = 114.6, SD = 15.5; CW mean belief = 118.2, SD = 14.2; $t[28] = -.67$, $p = .51$, two-tailed). Similarly, there was no significant difference between the belief scores of RW's disbelievers and CW's disbelievers (RW mean belief = 78.8, SD = 17.9; CW mean belief = 79.6, SD = 14.8; $t[28] = -.13$, $p = .90$, two-tailed). To identify whether believers or disbelievers were shifting their syllogisms performance, it was decided to look separately at the syllogisms scores of CW and RW's believers and disbelievers. If the psi-believers shifted their syllogisms scores, then we would expect CW's believers to significantly differ from RW's believers in their syllogisms scores. We would also expect to see no

significant difference between the syllogisms scores of CW's disbelievers and RW's disbelievers. As Table 8 shows, this is exactly what was found.

	CW	RW	t
Psi-believers	10.3 (N = 14, SD = 2.2)	12.6 (N = 16, SD = 3.4)	2.16 (.04)
Psi-disbelievers	12.3 (N = 16, SD = 4.7)	13.0 (N = 13, SD = 2.6)	.47 (.64)

Table 8. Mean syllogisms scores (N, SD in parentheses) for psi-believers and psi-disbelievers, by experimenter. The final column gives the results of t-tests comparing CW's believers' scores with RW's believers' scores, and CW's disbelievers' scores with RW's disbelievers' scores (two-tailed p-values in parentheses).

Conducting the same post hoc analysis for believers' and disbelievers' scores on the Ravens Matrices, the same pattern of scoring was found, suggesting a shift in believers' performance on the matrices task (see Table 9). CW's psi-believers had lower matrices scores than RW's believers. There was little difference between the matrices scores of the psi-disbelievers. This post hoc analysis gives some indication of an experimenter effect even for the Matrices task, though the planned correlational analyses did not show any evidence of an experimenter effect or of a belief-matrices performance correlation. Perhaps, with a small potential range of scoring possible for the matrices task, relative to the syllogisms task, the planned correlational analysis was insufficiently sensitive to detect any belief-matrices relationship. The post hoc analysis on the other hand seems to show a tendency for the same pattern of scoring as was found for the syllogisms task.

	CW	RW	t
Psi-believers	7.9 (N = 14, SD = 2.3)	9.2 (N = 16, SD = 1.0)	1.97 (.06)
Psi-disbelievers	9.3 (N = 15, SD = 1.8)	8.9 (N = 14, SD = 1.5)	-.65 (.52)

Table 9. Mean Ravens Matrices scores (N, SD in parentheses) for psi-believers and psi-disbelievers, by experimenter. The final column gives the results of t-tests comparing CW's believers' scores with RW's believers' scores, and CW's disbelievers' scores with RW's disbelievers' scores (two-tailed p-values in parentheses).

In sum, we have found clear evidence of an experimenter effect for the belief-syllogisms correlation, with CW finding a negative belief-syllogisms correlation and RW finding no correlation. Post hoc we have also found some evidence of a similar experimenter effect on the matrices task. Overall, then, we have provided evidence that different experimenters researching the relationship between belief in the paranormal and cognitive ability can produce different patterns of correlation. Our data indicate that it is the psi-believers who are responding differentially to the two experimenters. In attempting to relate our specific findings to the wider literature, two plausible scenarios then emerge:

1. In the general population there exists a negative belief-cognitive ability relationship, and CW's participants accurately represented that relationship. RW's participants on the other hand shifted their performance on the cognitive tasks, resulting in no belief-cognitive ability relationship. Given that RW's participants had slightly higher syllogisms and matrices scores, we may suggest that his believing participants were more motivated to perform well on the cognitive tasks, while his disbelieving

participants were unaffected. This pattern of scoring would remove the belief-cognitive ability correlation and would lead to slightly higher scoring overall on the cognitive tasks, which is what we see in RW's data. RW's position as a psi-skeptic is widely known so perhaps his believing participants felt relatively more motivated than his disbelieving participants to demonstrate their reasoning abilities. CW does not have a reputation as a strong psi-proponent, and ratings of her interaction with participants confirm that she appeared to be relatively neutral on the topic compared to RW. Her believing participants may then have felt little motivation to try harder on the cognitive tasks.

2. In the general population there is no belief-cognitive ability correlation, and RW's participants accurately reflected that position. For instance, it is possible that the existing literature showing a negative belief-cognitive ability correlation represents a reporting bias, particularly since many of these studies have been published by skeptics who might have the expectation that believers are cognitively "inferior" to disbelievers. Perhaps a number of studies showing null or even positive belief-cognitive ability correlations reside in these researchers' file-drawers. In this case, CW's believing participants seem to have shifted their performance downwards on the cognitive tasks, whilst the disbelieving participants were unaffected. This would lead to overall lower scoring for CW's participants on the cognitive tasks, and a negative belief-cognitive ability correlation, as was found for the syllogisms task. It is difficult to find a rationale for CW's believing participants to lower their performance on the cognitive tasks, given that CW is less skeptical about psi than RW. However, CW and RW vary on many other dimensions apart from attitude towards the paranormal, so possibly some other unidentified factor demotivated CW's participants.

We cannot go any further empirically to decide which of these two main scenarios (and there may be others that are less plausible) is valid. However, we regard the first as more parsimonious, due to RW's well-known position as a skeptic. Independent coding of the degree of skepticism shown by each experimenter in the initial

interaction with participants confirms that RW appeared to be more skeptical than CW. See Appendix 3 for additional group-level analyses.

One limitation of the present study is that both experimenters knew that it was a study of experimenter effects, and knew that they were being filmed during the session (Irwin, personal communication). To some extent, this may have limited the ecological validity of the study, as the experimenters' behaviour may have been altered by this knowledge, compared to a study where the experimenters were not consciously framing the study in terms of experimenter effects. However, in practice it is difficult to conduct such a study blind in the relatively small field of parapsychology, where the psi research track record of experimenters is already well-known and where the pairing of two experimenters with opposite beliefs would implicitly suggest the study was about experimenter effects. Also, for ethical reasons it is preferable to inform participants that the session will be filmed, in which case it would be difficult to keep the experimenter unaware of filming.

The second strand of this study investigated whether two experimenters with differing beliefs about psi would obtain different patterns of psi scoring from their respective participants. We predicted on the basis of past research that RW's participants would show less evidence of psi than CW's participants. This prediction was not confirmed, indeed there was a non-significant tendency for RW's participants to have greater psi scoring than CW's participants. Post hoc, we may note that RW appeared to have a bi-modal distribution of target ranks, with a relatively large number of rank 1 targets and rank 5 targets. This leads to RW's mean target rank being non-significantly different from CW's. An exact binomial analysis of RW's direct hits gives a significant result, ($Z = 1.95$, $p = .05$, 2-t). However, given that this analysis was not planned in advance, we should treat this finding with caution.

As there was some indication that RW obtained significantly more direct hits than chance, while CW's hit-rate was at chance, it was decided to examine the independent

ratings of the filmed psi task interviews. This might throw light on differences between the experimenters' ability to elicit psi-related material from participants. As described above, the recordings were rated for spontaneity of participants' utterances, quantity of images/ideas expressed, and amount of detail conveyed by participants. Table 10 below shows the descriptive statistics for these ratings for each experimenter. The only significant difference between the experimenters was that RW elicited more detail than CW (by a Mann-Whitney test, spontaneity $Z_{\text{corrected for ties}} = -1.48$, $p = .14$, two-tailed ; quantity $Z_{\text{corrected for ties}} = -.61$, $p = .54$, two-tailed; Detail $Z_{\text{corrected for ties}} = -2.71$, $p = .01$, two-tailed). In order to explore whether the additional detail elicited by RW relative to CW may have contributed to his larger number of hits, target rank was correlated with detail for RW. The correlation was close to zero ($r_{\text{scorrected}}$ for ties $= .02$, $p = .92$, two-tailed) which suggests that the additional detail elicited by RW did not seem to have been useful detail for identifying the target.

	Spontaneity	**Quantity**	**Detail**
CW	4.8 (1.6)	3.2 (1.3)	3.5 (1.7)
RW	4.2 (1.6)	3.4 (1.6)	4.7 (1.7)

Table 10. Mean ratings of participants' psi impressions (SD in parentheses) for CW and RW.

It is not clear why we failed to confirm the previous pattern of results obtained by Wiseman and Schlitz (1998, 1999). One plausible explanation is that although CW is undoubtedly more open to psi than RW and although CW devotes a relatively greater proportion of her research efforts towards testing the psi hypothesis than RW, her psi research track record has provided many fewer positive psi results than MS. This may lead CW to have lower expectations of success in her psi studies than MS. In this case, it is quite possible that MS, deliberately or unconsciously, may communicate more positive expectancies to her participants than CW. Relative to

MS, then, CW may resemble a "neutral" experimenter rather than being a strong psi-proponent. There is some support for this from the finding that CW was rated by an independent coder as "neutral" in ten out of twelve sessions. Another possibility is that CW simply has less psi ability than MS.

Many commentators have noted the inconsistent findings of the literature researching the link between belief in the paranormal and cognitive ability. Some have suggested that these inconsistencies may be due to experimenter effects. Our study is the first to provide evidence of an experimenter effect in investigating the belief-cognitive ability relationship. Our results do not support Irwin's specific rationale for the previous research findings (that skeptical experimenters may inhibit critically thinking participants' admission of their paranormal beliefs). Instead, exploratory analysis has suggested that in our study believing participants may have shifted their performance on the cognitive tasks. We cannot draw many conclusions as to whether it was the experimenters' differing attitudes towards psi that caused our pattern of results, because CW and RW vary on many dimensions apart from attitude towards the paranormal. Future research will be needed to uncover the mechanisms underlying this effect. The mechanisms are likely to be complex. For instance, it has been suggested that critical thinking ability may be differentially applied in different domains, and that attitudes and personality traits may be important factors in moderating the generalizability of critical thinking skills (Royalty, 1995). Research into interpersonal expectancy effects suggests that the experimenter's motivation for control and the participant's need for social approval may both be important moderating factors in the operation of such effects (Hazelrigg, Cooper & Strathman, 1991). In the meantime, it is vital that parapsychologists and psychologists researching this area realise that their participants' performance may be affected by the experimental context. The experimenter's own beliefs and idiosyncracies in interacting with participants may affect participants so as to elicit, or to obscure, a relationship between paranormal belief and cognitive ability.

References

Alcock, J.E., & Otis, L.P. (1980). Critical thinking and belief in the paranormal. Psychological Reports, 46, 479-482.

Blackmore, S., & Troscianko, T. (1985). Belief in the paranormal: Probability judgements, illusory control, and the "chance baseline shift". British Journal of Psychology, 76, 459-468.

Cooper, H., & Hazelrigg, P. (1988). Personality moderators of interpersonal expectancy effects: An integrative research review. Journal of Personality and Social Psychology, 55, 937-949.

Delanoy, D.L. (1988). Characteristics of successful free-response targets: Experimental findings and observations. Proceedings of the 31st Annual Convention of the Parapsychological Association, 230-246.

Delanoy, D.L., Watt, C.A., Morris, R.L. & Wiseman, R. (1993). A new methodology for free-response ESP testing outwith the laboratory: Findings from experienced participants. The Parapsychological Association 36th Annual Convention: Proceedings of presented papers (pp.204-221).

Edge, H.L., Morris, R.L., Palmer, J., & Rush, J.H. (1986). Foundations of Parapsychology: Exploring the boundaries of human capability. London: Routledge & Kegan Paul.

French, C.C. (1992). Factors underlying belief in the paranormal: Do sheep and goats think differently? The Psychologist: Bulletin of the British Psychological Society, 5, 295-299.

Harris, M.J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. Psychological Bulletin, 97, 363-386.

Hazelrigg, P., Cooper, H., & Strathman, A.J. (1991). Personality moderators of the experimenter expectancy effect: A reexamination of five hypotheses. Personality and Social Psychology Bulletin, 17, 569-579.

Irwin, H. (1991). Reasoning skills of paranormal believers. Journal of Parapsychology, 55(3), 281-300.

Irwin, H, (1999). An Introduction to Parapsychology 3rd ed. London: McFarland & Co.

Irwin, H. Personal email communication to CW, 19th November 2001.

Lange, R., Irwin, H.J., & Houran, J. (2000). Top-down purification of Tobacyk's Revised Paranormal Belief Scale. Personality and Individual Differences, 29, 131-156.

Luborsky, L., Diguer, L., Seligman, D.A., Rosenthal, R., Krause, E.D., Johnson, S., Halperin, G., Bishop, M., Berman, J.S. and Schweizer, E. (1999). The Researcher's Own Therapy Allegiances: A 'Wild Card' in Comparisons of Treatment Efficacy, Clinical Psychology: Science and Practice, 6(1), 95-106.

Merla-Ramos, M. (2000). Belief and reasoning: the effects of beliefs on syllogistic reasoning. Dissertation Abstracts International: Section B: The Sciences and Engineering. 61, 558.

Palmer, J. (1989a). Confronting the Experimenter Effect. Part 1.

Parapsychology Review, 20 (4), 1-4.

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

Palmer, J. (1997). The challenge of experimenter psi. European Journal of Parapsychology, 13, 110-125.

Rand Corporation (1955). A million random digits and 100,000 normal deviates. New York: Free Press.

Raven, J.C., Court, J.H., & Raven, J. (1985). A Manual for Raven's Progressive Matrices and Vocabulary Scales. London: H.K. Lewis.

Roberts, M., & Seager, P B. (1999). Predicting belief in paranormal phenomena: A comparison of conditional and probabilistic reasoning. Applied Cognitive Psychology, 13, 443-450.

Roe, C.A. (1999). Critical thinking and belief in the paranormal: A re-evaluation. British Journal of Psychology, 90, 85-98.

Rosenthal, R. (1976). Experimenter effects in Behavioual Research (enlarged ed.). New York: Irvington Press.

Rosenthal, R., & Rubin, D.B. (1978). Interpersonal expectancy effects: the first 345 studies. Behavioral and Brain Sciences, 3, 377-415.

Rosenthal, R. (1990). Replication in behavioural research. Journal of Social Behaviour and Personality, 5, 1-30.

Royalty, J. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. The Journal of Genetic Psychology, 156, 477-488.

Smith, M.D., Foster, C.L., & Stovin, G. (1998). Intelligence and paranormal belief: Examining the role of context. Journal of Parapsychology, 62, 65-78.

Solfvin, G.F., Kelly, E.F., & Burdick, D.S. (1978). Some new methods of analysis for preferential-ranking data. Journal of the American Society for Psychical Research, 72, 93-110.

Schmeidler, G.R. (1997). Psi-conducive experimenters and psi-permissive ones. European Journal of Parapsychology, 13, 83-94.

Targ, R., & Puthoff, H. (1977). Mind Reach: Scientists look at psychic ability. New York: Delta.

Tobacyk, J.J. (1988). A Revised Paranormal Belief Scale. Unpublished manuscript, Louisiana Tech University, Rushton, LA.

Watt, C.A. (1988). Characteristics of successful free-response targets: Theoretical considerations. Proceedings of the 31st Annual Convention of the Parapsychological Association, 247-263.

Watt, C.A. (1996). Knowing the unknown: Participants' insight in three forced-choice ESP studies. Journal of the American Society for Psychical Research, 90, 97-114.

Watt, C., Ravenscroft, J., & McDermott, Z. (1999). Exploring the limits of direct mental influence: Two studies comparing "blocking" and "co-operating" strategies. Journal for Scientific Exploration, 13, 515-535.

Wierzbicki, M. (1985). Reasoning errors and belief in the paranormal. Journal of Social Psychology, 125, 489-494.

Wiseman, R. (1997). Deception and Self-Deception: Investigating Psychics. Amherst: Prometheus.

Wiseman, R. & Greening, E. (1998). Remote viewing: An attempted replication. Proceedings of the 22nd International Conference of the Society for Psychical Research, 35.

Wiseman, R. & Schlitz, M. (1998). Experimenter effects and the remote detection of staring. Journal of Parapsychology, 61(3), 197-208.

Wiseman, R. & Schlitz, M. (1999). Replication of experimenter effects and the remote detection of staring. Proceedings of the 43rd Annual Convention of the Parapsychological Association. USA, 147-153.

Wiseman, R. & Watt, C. (1999). Rupert Sheldrake and the Objectivity of Science, Skeptical Inquirer, 23 (5), 61-62.

Appendix 1. Syllogisms Questionnaire.

Participant number_____

In this questionnaire you are given a series of reasoning problems. All problems have the same format. Each problem has two statements (i) and (ii) followed by a conclusion. For the sake of this exercise you are to proceed on the assumption that statements (i) and (ii) are true, even if you don't agree with them. Your task is to read the two statements and then determine if the conclusion *follows logically* from the statements, that is, if the conclusion is VALID or INVALID. Indicate your decision by circling either VALID or INVALID in the Response Column.

Note particularly that you are asked to judge if the argument from the pair of statements to the conclusion is sound. This is *not* simply a matter of whether the statements and/or the conclusion are true or false.

Consider the following example:

Statement (i): All birds can fly.
Statement (ii): An emu can not fly.

Conclusion: Therefore an emu is not a bird. VALID INVALID

Now, you might dispute the truth of one or more of the sentences in the above example, but irrespective of this, the task here is to judge if the concluding sentence is a proper deduction from the first two sentences. If you think the reasoning is sound you would circle VALID; if you think the reasoning is unsound you would circle INVALID.

The argument in the above example is indeed a VALID one, despite the conclusion being a false claim. That is, if the two statements are accepted, then the given conclusion is a legitimate deduction from them. You would therefore be expected to circle VALID for this example.

Try to think clearly about these problems, but do not spend too much time on any one of them.

		Response Column	
1.	(i) If a book is well written it is worth reading. (ii) This book is well written. Therefore this book is worth reading.	VALID	INVALID
2.	(i) All wise people behave in an ethical manner. (ii) This person is wise. Therefore this person behaves in an ethical manner.	VALID	INVALID
3.	(i) All insects have six legs. (ii) This creature has six legs. Therefore this creature is an insect	VALID	INVALID
4.	(i) If a cow is properly fed it will yield milk of a high quality. (ii) This cow yields milk of a high quality. Therefore this cow is properly fed.	VALID	INVALID
5.	(i) All University of Hertfordshire students are members of the Students Union. (ii) A student at the Hatfield Campus is a University of Hertfordshire student. Therefore a student at the Hatfield Campus is a member of the Students Union.	VALID	INVALID
6.	(i) If a dog is treated kindly it is loyal. (ii) This dog is not loyal. Therefore this dog is not treated kindly.	VALID	INVALID

7. (i) If furniture is skilfully made it will last for centuries.
(ii) This piece of furniture lasted only six months.

Therefore this piece of furniture was not skilfully made. VALID INVALID

8. (i) If businessmen are completely ruthless they will make a large profit.
(ii) This businessman is not completely ruthless.

Therefore this businessman will not make a large profit. VALID INVALID

9. (i) All poisonous substances taste bitter.
(ii) Strychnine tastes bitter.

Therefore strychnine is a poisonous substance. VALID INVALID

10. (i) If people are your true friends you can trust them with a secret.
(ii) You can trust this person with a secret.

Therefore this person is a true friend of yours. VALID INVALID

11. (i) If citizens of a country are free to criticise their government the government is able to respond to the real needs of the people.
(ii) Canadians are free to criticise their government.

Therefore the Canadian government is able to respond to the real needs of its people. VALID INVALID

12. (i) If people have markedly defective vision they wear glasses.
(ii) This student does not have markedly defective vision.

Therefore this student does not wear glasses. VALID INVALID

13. (i) If a cow is properly fed it will yield milk of a high quality.
(ii) This cow is properly fed.

Therefore this cow will yield milk of a high quality. VALID INVALID

14. (i) All wise people behave in an ethical manner.
(ii) This person behaves in an ethical manner.

Therefore this person is wise. VALID INVALID

15. (i) If businessmen are completely ruthless they will make a large profit.
(ii) This businessman does not make a large profit.

Therefore this businessman is not completely ruthless. VALID INVALID

16. (i) All insects have six legs.
(ii) This creature is an insect.

Therefore this creature has six legs. VALID INVALID

17. (i) If a dog is treated kindly it is loyal.
(ii) This dog is not treated kindly.

Therefore this dog is not loyal. VALID INVALID

18. (i) If the citizens of a country are free to criticise their government the government is able to respond to the real needs of the people.
(ii) The Canadian government is able to respond to the real needs of its people.

Therefore Canadians are free to criticise their government. VALID INVALID

19. (i) If a book is well written it is worth reading.
(ii) This book is worth reading.

Therefore this book is well written. VALID INVALID

20. (i) If people are your true friends you can trust them with a secret.
(ii) This person is a true friend of yours.

Therefore you can trust this person with a secret. VALID INVALID

21. (i) All University of Hertfordshire students are members of the Students Union.
(ii) A student at the Hatfield campus is a member of the Students Union.

Therefore a student at the Hatfield campus is a VALID INVALID

University of Hertfordshire student.

22. (i) All poisonous substances taste bitter.

(ii) Strychnine is a poisonous substance.

Therefore strychnine tastes bitter. VALID INVALID

23. (i) If furniture is skilfully made it will last for centuries.

(ii) This piece of furniture is not skilfully made.

Therefore this piece of furniture will not last for centuries. VALID INVALID

24. (i) If people have markedly defective vision they wear glasses

(ii) This student does not wear glasses.

Therefore this student does not have markedly defective vision. VALID INVALID

Please ensure that you have answered every question.

Appendix 2.

Overall distribution of target ranks excluding two sessions in which the wrong target was played to the participant (N = 58). Obtained sum-of-ranks is 165, MCE sum-of-ranks is 174. This gives a non-significant Z = .07 (using the convention of a positive sign for the Z where the data is in the psi-hitting direction).

	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**
Actual	17	10	9	9	13
MCE	11.6	11.6	11.6	11.6	11.6

Distribution of target ranks for CW and RW, excluding two sessions in which wrong target was played to the participant (N = 58). The mean target rank for RW is 2.7 (SD = 1.6), and CW's mean target rank is exactly at chance (mean = 3.0, SD = 1.5). This difference is not significant on a Mann-Whitney test ($Z_{\text{corrected for ties}} = -.90$, p = .37, two-tailed).

	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**
CW	6	6	4	6	6
RW	11	4	5	3	7

Appendix 3

Group-level analyses.

By independently coding the film of the experimenters' initial interaction with participants, we have seen how the experimenters differed in the degree of psi belief or skepticism they portrayed. This served as a manipulation check, as reported in the main paper. Secondly, it may be informative to correlate the experimenter's degree of skepticism with each session's syllogism scores (we focus here on syllogisms because the strongest belief-cognitive ability correlation was found for this measure), and with each session's belief scores.

Group belief scores and group syllogisms scores.

As participants were tested in groups of one to three people, a mean syllogisms score and a mean belief score was calculated for each group of participants. This gave 12 sets of scores for RW's groups and 12 for CW's groups. Interestingly, even at the group level, with N reduced from 30 to 12, CW's groups showed a marginally significant negative correlation between overall belief in the paranormal and syllogisms performance ($r = -.515$, $p = .09$, two-tailed). This indicates a relatively robust effect that exists at the group level for CW. RW's groups continued to show a non-significant correlation ($r = .375$, $p = .23$).

Experimenter skepticism rating and group syllogisms scoring.

The skepticism ratings for CW show that her apparent psi belief varied little between sessions, and therefore the correlation between her skepticism rating and syllogisms scores is likely to be unreliable ($r_{s,\text{ corrected for ties}} = .533$, $p = .08$, two-tailed). When RW's skepticism ratings are correlated with his groups' syllogisms scores, a marginally significant negative correlation is found ($r_{s,\text{ corrected for ties}} = -.555$, $p = .06$, two-tailed). Therefore these correlations suggest an interaction between the experimenter's skepticism rating and group syllogisms scores: when RW is less skeptical, his group syllogisms scores are larger; when CW is less skeptical, her group

syllogisms scores are smaller. This interaction is depicted pictorially in figures 1 and 2 below (reminder: rating 1 = proponent, rating 7 = skeptic).

Experimenter skepticism rating and group belief scoring.

The main paper argued that our data suggested that participants' syllogisms scoring was being affected by the experimenter, rather than participants' paranormal belief scores. This argument is supported by additional group-level analyses, where experimenter skepticism rating is correlated with the groups' belief scores. Neither experimenter obtained a significant correlation between experimenter skepticism rating and group belief scores (CW $r_{s,\text{ corrected for ties}} = -.326$, $p = .28$, two-tailed; RW $r_{s,\text{ corrected for ties}} = .092$, $p = .76$, two-tailed).

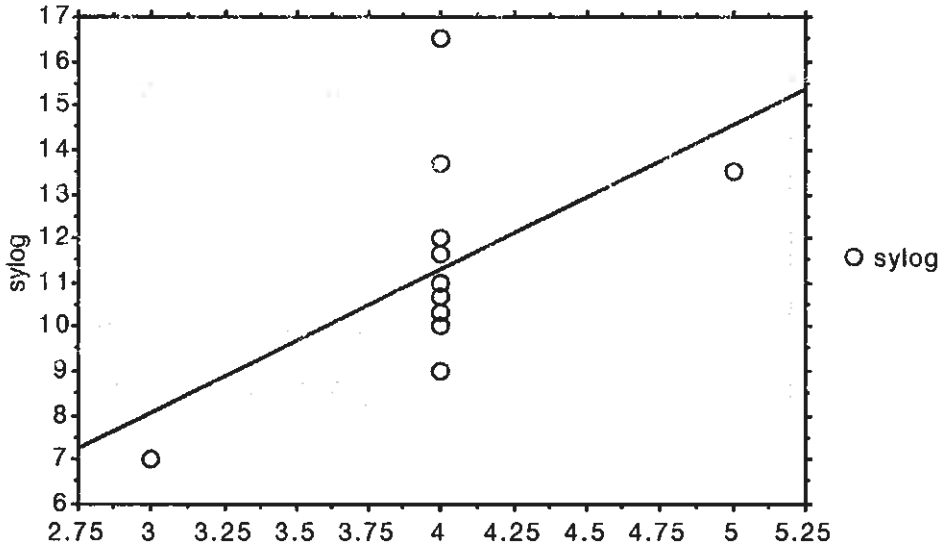

Figure 1: CW's skepticism ratings vs group syllogisms scores.

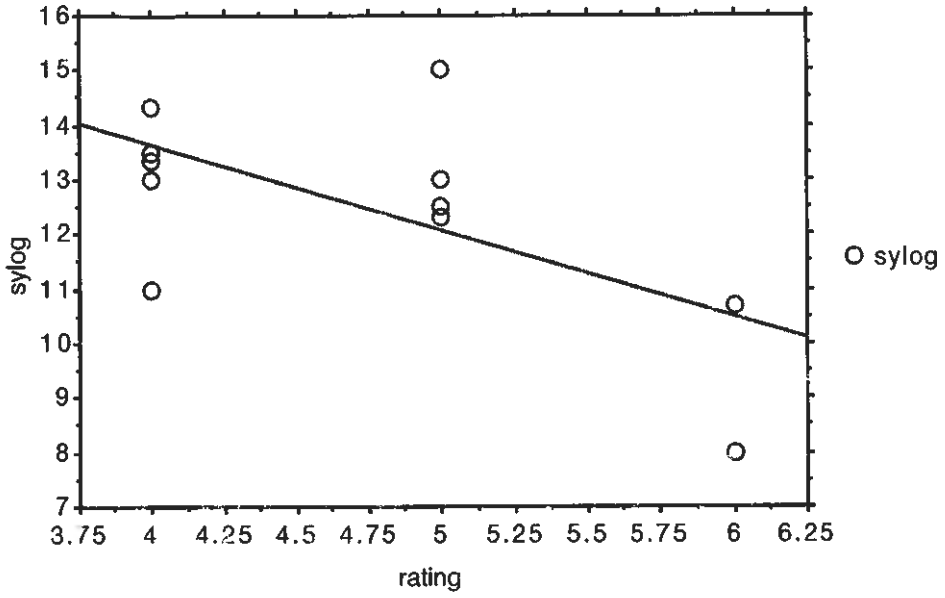

Figure 2: RW's skepticism ratings vs group syllogisms scores.