RESEARCH ARTICLE

Two Replication Studies of a Time-Reversed (Psi) Priming Task and the Role of Expectancy in Reaction Times

Marilyn Schlitz
Sofia University, Palto Alto, California, USA
mjmschlitz@gmail.com

Daryl Bem
Cornell University, Ithaca, New York, USA

David Marcusson-Clavertz
Lund University, Lund, Sweden

Etzel Cardeña
Lund University, Lund, Sweden

Jennifer Lyke
Stockton University, Galloway, New Jersey, USA

Raman Grover
Vancouver, British Columbia, Canada

Susan Blackmore
University of Plymouth, Plymouth, UK

Patrizio Tressoldi
University of Padova, Padova, Italy

Serena Roney-Dougal
Psi Research Center, Glastonbury, UK

Dick Bierman
University of Groningen, The Netherlands

Jacob Jolij
University of Groningen, The Netherlands

Eva Lobach
University of Amsterdam, The Netherlands

Glenn Hartelius
California Institute of Integral Studies, San Francisco, California, USA

Arnaud Delorme
University of California, San Diego, California, USA
Institute of Noetic Sciences, Petaluma, California, USA

Abstract—Two experiments involving an international collaboration of experimenters sought to replicate and extend a previously published psi experiment on precognition by Daryl Bem that has been the focus of extensive research. The experiment reverses the usual cause–effect sequence of a standard psychology experiment using priming and reaction times. The preregistered confirmatory hypothesis is that response times to incongruent stimuli will be longer than response times to congruent stimuli even though the prime has not yet appeared when the participant records their judgments. The confirmatory hypothesis for Experiment 1 was not supported. Exploratory analyses indicated that those participants who completed the English-language version rather than a translation showed a significant effect, as was the case in the original study; no significant departure from chance was found in data involving non-English translations. Experiment 2 sought to enhance the predicted effect by having each participant read either a pro-psi or an anti-psi statement at the beginning of the experiment to test the hypothesis that a pro-psi statement would produce a larger effect than an anti-psi statement. The results did not support the primary psi hypothesis and there was no effect in the English-language sample. However, there was mixed support for the effect of the psi statement on performance; those participants who received the pro-psi statement had a greater psi score than those who received the anti-psi statement. As in the original experiment, neither the experimenters' nor participants' beliefs were consistently associated with the dependent measure. In sum, the pre-registered confirmatory hypotheses were not supported. The importance of the personality variable Sensation Seeking, a component of extraversion, as a correlate of psi performance is discussed as are the challenges and implications for international collaborations and replication in controversial science.

Keywords: priming; expectancy effect; retrocausation; consciousness; sociology; precognition; psi; replication

Psi research involves the study of purported anomalous mental phenomena, including telepathy, clairvoyance, precognition, and psychokinesis (mind over matter). It is an area of controversial science that began in the late 1800s and continues today. Although several meta-analytic reviews demonstrate an overall psi effect (for a review, see Cardeña, 2018), meta-analyses may suffer from publication biases and selective reporting. For instance, a recent comparison of effect sizes in meta-analyses and pre-registered replication attempts across

15 domains of psychology in general indicated that the former showed almost three times as large effects (Kvarven et al., 2020). Thus, this project sought to address these challenges by examining whether independent investigators can replicate reportedly successful psi experiments using pre-registered analyses.

In 2011, Bem published results from a series of precognition experiments in the *Journal of Personality and Social Psychology*. Using a variety of protocols, his nine experiments tested for possible retroactive (i.e. "precognitive") influences of well-established psychological effects by "time-reversing" the stimulus and response: On each trial, a participant's response was recorded *before* the purportedly causal stimulus was presented. Bem reported statistically significant results in eight of the nine experiments, with a statistically significant mean effect size (d) of 0.22 (Stouffer's $z = 6.66$, $p = 1.34 \times 10^{-11}$). Critics argued that the analyses were partly exploratory (Wagenmakers et al., 2011) and low-powered (Schimmack, 2012), which may result in false-positives. To encourage independent replications, Bem made all his experimental materials and instructions available to other investigators. By 2016, Bem et al. were able to report a meta-analysis of 90 such experiments from 33 laboratories in 14 countries. This yielded an overall effect size greater than 6 sigma, with a Bayes Factor of 3.8×10^9, greatly exceeding the criterion value of 100 for "decisive evidence" in support of the experimental hypothesis. With Bem's original studies excluded, the effect remained significant albeit small, ES = 0.06, $z = 4.16$, $p = 1.1 \times 10^{-5}$, BF = 3.85.

An important variable in determining the success or failure of experimental hypotheses is the experimenter's orientation toward the phenomenon under investigation (Collins & Pinch, 1979). In mainstream psychology, Rosenthal has demonstrated experimenter expectation effects in more than 300 studies, including studies in classroom and clinical settings (Rosenthal, 1978). Experimenter effects have also been observed in psi research for more than 70 years (Pratt et al., 1940), with Palmer and Millar (2015) suggesting that experimenter effects are important or even crucial determinants of outcomes in psi research. Krippner (1978) has summarized findings showing differences among experimenters (Honorton et al., 1975), data collectors (Johnson et al., 1972), reciprocal attitudes between experimenter and participant

(Nash, 1968), and differences across time by the same experimenter (Rivers, 1950). Although experimenter effects are usually attributed to sensory cues, some researchers have suggested that some may be psi-mediated (e.g., Honorton, 1978; Kennedy & Taddonio, 1976; Thouless, 1976; White, 1977). For example, participants did better at guessing psi targets prepared by a psi proponent than on those prepared by a psi skeptic (West & Fisk, 1953).

There is also evidence suggesting that an experimenter can remotely influence a participant's responses through the mediation of psi. For example, Schlitz and Braud (1997) reported that experimenters influenced a participant's electrodermal activity from a distance. Using this protocol, Schlitz and psi-skeptic Richard Wiseman collaborated in three attempted replications using the same participant pool and procedures. Schlitz obtained significant psi effects in two of the three experiments, but Wiseman failed to generate results that allowed for the rejection of the null hypothesis (Watt et al., 2005). Roe et al. (2006) also studied the effect of the experimenter on outcomes of two psi experiments and found that the more experienced experimenter obtained better results.

Each of the two experiments reported here sought to replicate Bem's (2011) two experiments on retroactive priming and to examine the possible effects of the experimenters' and participants' beliefs about psi on the outcome of the experiment in which they were participating. On each trial of a standard (i.e. non-psi) priming task, a pleasant or unpleasant word (the "prime") is briefly shown on a computer screen followed immediately by a pleasant or unpleasant picture drawn from the standard International Affective Picture System (IAPS) (Lang et al., 1993). Trials on which the image and the priming word have different valences (one pleasant and one unpleasant) are termed "Incongruent Trials"; trials on which the picture and the priming word share a common valence (both pleasant or both unpleasant) are termed "Congruent Trials." The typical finding is that participants respond more slowly on Incongruent trials than on Congruent ones.

In Bem's "time-reversed" psi version of the experiment, the presumed cause–effect sequence is reversed so that the prime is not flashed until *after* the participant has already recorded his or her judgment of the picture's valence. The experimental hypothesis

remains the same as in the standard non–time-reversed experiment: Response times will be longer on trials with Incongruent prime/picture pairs than on trials with Congruent prime/picture pairs. Both of Bem's time-reversed priming experiments were successful (Bem, 2011), and the followup meta-analysis of 15 precognitive priming experiments supported the hypothesis with an effect size (d) of 0.11, $p = 0.003$ (Bem et al., 2016).

On each trial of the procedure, two potential primes are pre-designated for the picture, one pleasant and one unpleasant. Immediately after the participant records his or her judgment of the picture's valence, the computer randomly selects one of the two words to serve as the priming word and flashes it briefly on the screen. This procedure thus provides a genuine sampling-with-replacement or "open deck" procedure for determining whether a trial will be congruent or incongruent. Accordingly, the probability that the trial will be congruent or incongruent remains constant at 0.5 across all trials. As a result, there is no (non-psi) way for a participant to anticipate the kind of trial currently on the screen.

In his original psi article, Bem (2011) noted that the personality trait of extraversion has been frequently reported over the years to be an individual-difference correlate of psi performance, with extraverts achieving higher psi scores than introverts. An analysis of 60 independent psi experiments published between 1945 and 1983 revealed a small but reliable correlation between extraversion and psi performances, $r = 0.09$, $z = 4.63$, $p = 0.000004$ (Honorton et al., 1992). And the correlation was observed again in a later set of telepathy studies conducted in Honorton's own laboratory, $r = 0.18$, $t(216) = 2.67$, $p = 0.004$ (Bem & Honorton, 1994).

The component of extraversion that appears to underlie this correlation is the extravert's susceptibility to boredom and a tendency to seek out stimulation. Specifically, Eysenck (1966) attributed the positive correlation between extraversion and psi to the observation that extraverts "are more susceptible to monotony . . . [and] respond more favourably to novel stimuli" (p. 59). Sensation Seeking is one of the six factors of extraversion on the Revised NEO Personality Inventory (Costa & McCrae, 1992), and Zuckerman's Sensation Seeking Scale (1974), which contains a subscale of Boredom Susceptibility, is

moderately correlated with overall extraversion, $r = 0.47$, $p < 0.01$ (Farley & Farley, 1967).

To assess Stimulus Seeking as a correlate of psi performance in seven of his nine "Feeling the Future" experiments, Bem constructed a scale comprising the two statements: (a) "I am easily bored," and (b) "I often enjoy seeing movies I've seen before" (reverse-scored). Responses were recorded on 5-point scales that ranged from *Very Untrue* to *Very True* and averaged into a single score ranging from 1 to 5. Stimulus Seeking was significantly correlated with psi performance in five of the seven experiments. The mean effect size was 0.43. Both experiments reported here involved three levels of participants: (a) Professors and other Investigators who recruited student experimenters and were invited to serve as participants themselves, (b) Student experimenters who received standardized training in the experimental procedure, and (c) Participants who engaged in the psi task. Investigators who conducted the experiment in a university setting and obtained their own Institutional Review Board (IRB) approvals were offered the option of co-authorship on the final report.

The two experiments were pre-registered with the Koestler Parapsychology Unit at:
http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_Registry_1007.pdf
and
http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_registry_1016.pdf
The pre-registered study design of Experiment 1 called for 32 experimenters who would test 16 participants each for a total of 512 participants. Drawing on a global professional network of teachers and other colleagues, 16 professors and other investigators agreed to participate. Of these, four dropped out because of time constraints and other issues. The remaining 12 recruited a total of 34 experimenters. The experimenters were selected based on their interest in the studies, but not on their beliefs in psi. As planned, the first 32 experimenters who submitted complete datasets for each study were included in the analysis (the two other experimenters did not return all the necessary datasets). The script for both studies is included in the **Appendix**.

Experiment 1: Retroactive Priming as a Function of Psi Experiences and Beliefs in Psi

EXPERIMENT 1 METHODS

The procedure was identical to Bem's (2011) original experiments for retroactive priming. Both experimenters and participants were assessed for their beliefs in psi phenomena, and experimenters were also assessed for their belief that the experiment would "produce evidence for precognition." As in the original studies, participants were informed ahead of time that the experiment would test for extrasensory perception (ESP). After they responded to the belief questions, participants went through a 3-minute relaxation procedure and then began the retroactive priming task.

Experiment 1 investigated three pre-registered hypotheses: one confirmatory and two exploratory. The confirmatory hypothesis was that (a) The previous effect reported by Bem (2011) would be successfully replicated: Response times (RT) on trials with Incongruent picture/prime pairs would be greater than RT on trials with Congruent picture/prime pairs. The two exploratory hypotheses were: (b) The anomalous RT effects would be greater for experimenters with positive beliefs about psi and more psi experiences than for experimenters with negative beliefs about psi and fewer psi experiences, and (c) The anomalous RT effects would be greater for participants with positive beliefs about psi and more psi experiences than for those with negative beliefs about psi and fewer psi experiences. The study was not powered for significance on the two secondary hypotheses but sought to identify a trend.

The main dependent variable of analysis in both Experiments was a participant's RT score, defined as their mean response time on Incongruent trials minus their mean response time on Congruent trials—with the following arithmetic modification: RT measurements are not normally distributed but are positively skewed with a lower bound of 0. Accordingly, it is routine practice in experiments using RT as the dependent variable to transform each raw RT measurement using either a reciprocal transformation (1/RT) or a log transformation log(RT), and to define outliers that are too short or too long using the

same transformation. To examine the robustness of the results we examined two data transformations (log and inverse) and two cutoffs for maximum RTs (2,500 ms and 1,500 ms). All trials in which RTs were below 250 ms or above the maximum (i.e. 2,500 ms or 1,500 ms) were excluded.

Mean RT scores greater than 0 imply that a participant's RT on Incongruent trials were longer than RT on Congruent trials, a confirmation of the psi hypothesis; RT scores equal to or less than 0 denote disconfirming instances of the hypothesis. In total, 32 experimenters and 512 participants completed the test. In contrast to the pre-registration, we analyzed experimenter and participant effects with a two-level mixed model with random intercept to account for the multilevel structure of the data as participants were nested within experimenters (but similar results were obtained with the ANOVA). In addition, the mixed model analysis allows for examining the proportion of variance in the retroactive priming effect at the experimenter level. Experimenters' and participants' beliefs and experiences of psi were categorized as high, medium, or low according to 33rd percentiles, as pre-registered. The analyses were performed in JAMOVI 1.2.27. The pre-registration included bootstrapping in Experiment 2, but because bootstrapping and regular parametric methods yielded similar results and JAMOVI did not include bootstrapping for these analyses, we report the parametric analyses.

EXPERIMENT 1 RESULTS

Summary of Data

In Experiment 1, we analyzed the data of 32 experimenters and 511 participants (languages; n_{Dutch} = 233, $n_{English}$ = 189, n_{French} = 48, $n_{Italian}$ = 25, $n_{Bulgarian}$ = 16). We first describe their self-report data. In both groups, the typical response to the question of whether ESP exists was that it "probably does," but there was a wide range of beliefs. Specifically, among the experimenters, 23% reported that they believe ESP definitely exists, 45% that it probably exists, 16% that they do not know, 13% that it probably does not exist, and 3% that it definitely does not exist (one experimenter did not answer the questionnaire). Among the participants, 18% reported believing that ESP definitely exists, 41%

that it probably does, 17% that they do not know, 14% that it probably does not exist, and 9% that it definitely does not exist. On average, experimenters and participants did not significantly differ in ESP belief, $t(540) = 1.15$, $p = .25$.

As for the practice of mental discipline (e.g., meditation, hypnosis), 29% of the experimenters reported having had regular practice, 29% reported occasional practice, and 42% reported practicing it only a few times or never. Among the participants, 18% reported having had regular practice, 18% reported occasional practice, and 64% reported having practiced it only a few times or never. On average, experimenters reported greater practice of mental discipline than participants, $t(540) = 2.47$, $p = .01$.

We next describe the behavioral data from the retroactive priming task. Participants accurately identified the images as "pleasant" or "unpleasant" in 92% of the trials. Seventeen participants (3%) had an error rate at 25% or higher. The data from these individuals were excluded from further analysis, as pre-registered. Figure 1 shows the raw and log transformed RT averaged for each participant and congruency condition using a 2,500 ms cutoff. As shown, skewness was moderately positive for *raw* data (0.89 for the congruent condition and 0.82 for the incongruent), but small for the *log* transformed data (0.34 for congruent and 0.30 for incongruent, respectively). The *inverse* transformed data also showed small skewness (0.22 and 0.21 for congruent and incongruent, respectively). Analogously, the data based on the 1,500 ms cutoff had absolute skewness values below 0.47. To summarize, the data transformations resulted in largely symmetric distributions as illustrated in Figure 1.

Confirmatory Analyses

The psi hypothesis states that RT will be longer for trials with incongruent than for congruent stimuli. The primary tests of this hypothesis are shown in Table 1. Neither of the four *t*-tests of the primary hypothesis yielded significant results (i.e. no significant retroactive priming effect). Thus, the null hypothesis could not be rejected. To supplement the primary analyses, we also performed binomial tests to examine whether the proportions of participants who scored positively
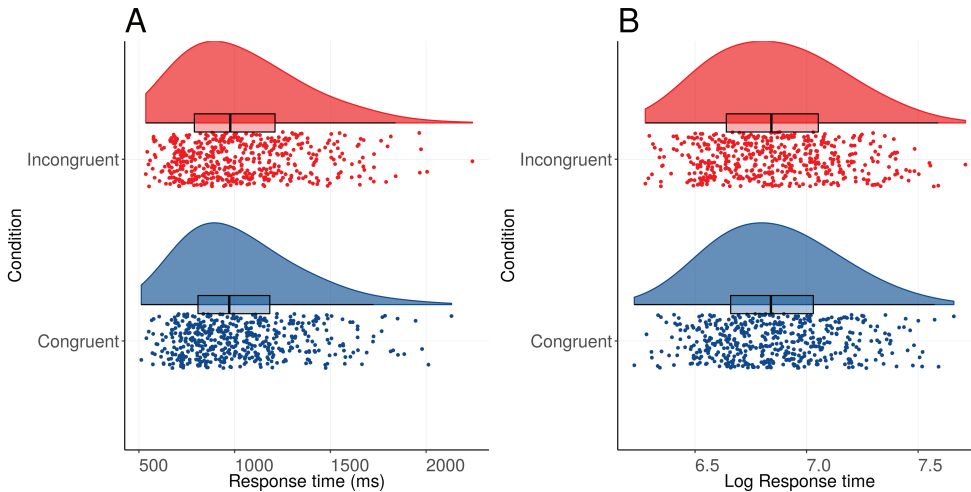
Figure 1. Illustration of RT distributions for (A) raw and (B) log transformed data in the retroactive priming task (Experiment 1). The psi hypothesis states that RT should be slower in the incongruent condition. Each dot represents the average score of a participant for that condition (jittered). The lower and upper hinge on the boxplots indicate the first and third quartile with the median in between. The whisker lines extend to the most extreme values but no further than 1.5 × the interquartile range from the hinge. The raincloud shapes represent distribution estimates based on Kernel density probability functions. Individuals with greater accuracy than .75 and trials with RT above 250 ms and below 2,500 ms are included in this analysis.

were greater than chance (mean chance expectation = 50%). This was done for each data transformation and cutoff. Consistent with the primary analyses, there was no support for the retroactive priming hypothesis in the full sample ($ps > .30$ for all four outcomes).

Exploratory Analyses

The two pre-registered exploratory hypotheses were not supported though they were not powered for significance. That is, RT differences for experimenters and participants did not depend on whether they had low, medium, or high psi beliefs/experiences ($Fs < 1.2$, $ps > .30$ across the four psi outcomes). It should be noted that Bem's (2011) original experiments also failed to find such associations. In addition, our mixed model analysis indicated that there was almost 0 variance at the experimenter level ($ICCs$ [intraclass correlation coefficients] $< 1.88 \times 10^{-15}$ across the four

TABLE 1
**One-Sample *t*-Tests from Experiment 1 (one-tailed, positive *t*-scores
reflect greater than 0 retroactive priming difference scores)**

Data	Cutoff	Confirmatory All languages			Exploratory English Language			Non-English Language		
		t (df)	*p*	*d*	*t (df)*	*p*	*d*	*t (df)*	*p*	*d*
Log(RT)	2,500	*t*(492) = 0.76	.22	0.03	*t*(177) = 2.08	.02*	0.16	*t*(314) = −0.66	.74	−0.04
	1,500	*t*(489) = 0.24	.40	0.01	*t*(177) = 1.23	.11	0.09	*t*(311) = −0.60	.73	−0.03
1/RT [a]	2,500	*t*(492) = 0.49	.31	0.02	*t*(177) = 1.99	.02*	0.15	*t*(314) = −0.98	.84	−0.06
	1,500	*t*(489) = 0.25	.40	0.01	*t*(177) = 1.49	.07	0.11	*t*(311) = −0.80	.79	−0.05

[a] The *t*-test and *d*-statistics for the inverse transformation were reversed back so that the psi hypothesis would have the mean be *greater than* 0.
* *p* < .05

outcomes), indicating that there was virtually no systematic between-experimenter variance in retroactive priming scores.

An exploratory, not pre-registered, analysis indicated a significant retroactive priming effect for those completing the English-language version (*t*-tests, see Table 1). This effect was observed with both log and inverse transformations of the data, but only with the more liberal 2,500 ms cutoff. Those completing the non-English versions scored non-significantly in the opposite direction (see Table 1). As for the supplementary binomial analyses, 56.2% of the English-language sample had a positive log transformed difference score regardless of whether 2,500 or 1,500 ms cutoffs were used (*p*s = .06), whereas for the inverse transformation the proportions of positive psi scores were 54.5% (*p* = .13) and 53.9% (*p* = .17) for the 2,500 ms and 1,500 ms, respectively. For the non-English samples the proportions of positive scores were lower than 48% across all data transformations (*p*s > .80). To summarize, the exploratory analyses indicated an effect among those performing English-language versions using the 2,500 ms cutoff and continuous scores.

As the exploratory psi analyses indicated an effect only in the English-language sample, we compared that sample to the non-English samples on the five predictor variables. The English-language sample reported greater practice of *mental discipline*, t (509) = 3.34 p < .001, d = 0.31; *belief in ESP*, t (509) = 4.26, p < .001, d = 0.39; *experience of ESP*, t (509) = 4.06, p < .001, d = 0.37; and *easily getting bored*, t (509) = 4.38 p < .001, d = 0.40, but the samples did not differ in the extent they *enjoy watching the same movies again*, t (509) = −1.48 p = .14, d = 0.14.

We correlated these five predictors, as well as Bem's (2011) two-item Stimulus-Seeking scale, with the four retroactive priming outcomes (log and inverse data with 2,500 and 1,500 ms cutoff), that is, performing 24 analyses (see Table 2). With unadjusted p-values and one-tailed tests, the only predictors that significantly predicted retroactive priming outcome was the variable *getting easily bored* (significant for both data transformations but only the 1,500 ms cutoff) and the Stimulus-Seeking scale (but only for the 1,500 ms log score). However, those p-values would not remain significant with a Bonferroni correction (e.g., dividing the alpha value by 6 because of the six predictors). Nevertheless, for exploratory purposes a followup analysis was performed on those completing the English-language version. This analysis indicated that easily getting bored significantly correlated with the retroactive priming effect across all four outcomes (rs between .19 and .22, p-values < .005). This variable did not correlate with the retroactive priming effect among those completing the non-English version (ps > .42). Similar patterns were obtained with the Stimulus-Seeking scale (rs between .16 and .22, ps below .02 in the English sample).

In summary, those completing the English-language version reported greater ESP belief and experience, practice of mental discipline, and more easily getting bored. With adjustments of alpha value for multiple analyses, none of the six predictors correlated with the retroactive priming effect in the full sample, although the English-language sample exhibited a significant correlation between the retroactive priming effect and easily getting bored (and Stimulus Seeking more broadly).

TABLE 2
Pearson _r_ Coefficients (_p_-values in parentheses) from Exploratory
Correlational Analyses with Self-Reported Predictor Variables and
Retroactive Priming Outcome (one-tailed, unadjusted _p_-values)

	Log RT		1/RT (reversed)	
Predictor	**2500 ms**	**1500 ms**	**2500 ms**	**1500 ms**
ESP belief	.05 (.13)	.00 (.49)	.04 (.20)	.00 (.50)
Practice mental discipline	.04 (.19)	−.00 (.52)	.03 (.24)	.00 (.50)
ESP experience	.02 (.35)	−.02 (.65)	.01 (.44)	−.02 (.66)
Easily getting bored	.05 (.16)	.10 (.02*)	.06 (.09)	.09 (.02*)
Enjoy watching the same movies again	−.04 (.21)	−0.01 (.44)	−.02 (.33)	−.01 (.45)
Stimulus Seeking (2 items) [a]	.06 (.10)	.07 (.05*)	0.06 (.10)	.07 (.06)

* $p < .05$
[a] Stimulus seeking is the mean of easily getting bored and enjoy watching the same movies again (reversed). The correlation coefficient is multiplied with –1 for the inverse transformation so that greater positive scores reflect greater retroactive priming effect across all four outcomes. Only individuals with mean accuracy > .75 are included ($n = 493$ for 2,500 ms cutoff and $n = 490$ for 1,500 ms cutoff).

Experiment 2: Retroactive Priming Effects as a Function of Reading Pro-Psi or Anti-Psi Arguments

Experiment 2 was identical to Experiment 1 except that participants read a (genuine) pro-psi or anti-psi statement before beginning the experimental trials. The pre-registered study design of Experiment 2 called for 32 experimenters who would test 20 participants each for a total of 640 participants. The confirmatory hypothesis was that mean RT would be longer for trials with Incongruent prime/picture pairs than for trials with Congruent pairs. The exploratory hypotheses were that the predicted RT effects would depend on: (a) experimenters' psi beliefs and experiences, (b) participants' psi beliefs and experiences, and (c) the interaction between experimenter's and participant's psi beliefs and experiences. In addition to these pre-registered hypotheses, we also explored if sessions with a positive prompt regarding psi phenomena would have greater psi effects than sessions with a negative prompt.

EXPERIMENT 2 METHODS

Prior to each session, subjects were randomly exposed to one of two prompts: pro-psi or anti-psi. These prompts were as follows.

Pro-Psi Introduction

Comment on psi (ESP) by Rupert Sheldrake, Ph.D., a biologist and author of more than 80 scientific papers and ten books. He was among the top 100 Global Thought Leaders for 2013, as ranked by the Duttweiler Institute in Zurich, Switzerland's leading think tank.

> Telepathy, ESP, and psychic/psi phenomena in general are real and backed up by convincing evidence; their investigation deserves to be part of science . . . I take seriously research within parapsychology. I think there is good evidence for precognitive dreams, and also for presentiment, whereby an emotional arousal can have a physiological arousing effect five or six seconds in advance.

Anti-Psi Introduction

Comment on psi (ESP) by Michael Shermer, Ph.D., the Founding Publisher of *Skeptic Magazine*, a monthly columnist for *Scientific American*, a regular contributor to Time.com, and Presidential Fellow at Chapman University.

> . . . a meta-analysis of . . . [psi] experiments found no evidence for psi, concluding that psi data are non-replicable, a fatal flaw in scientific research. In general, over the course of a century of research on psi, the tighter the controls on the experimental conditions, the weaker the psi effects seem to become, until they disappear entirely. This is a very strong indicator that ESP is not real.

EXPERIMENT 2 RESULTS

Summary of Data

In Experiment 2, we analyzed the data of 30 experimenters and 586 participants (languages: $n_{Dutch} = 409$, $n_{English} = 117$, $n_{German} = 42$, $n_{Swedish} = 18$). Mean accuracy on judging unpleasant and pleasant images was 92% (as in Experiment 1). Twenty-two participants (4%) were excluded

TABLE 3
One-Sample *t*-Tests from Experiment 2 (one-tailed, positive *t*-scores
reflect greater than 0 retroactive priming difference scores)

Data	Cutoff	Confirmatory All Languages			Exploratory English Language			Non-English Language		
		t (*df*)	*p*	*d*	*t* (*df*)	*p*	*d*	*t* (*df*)	*p*	*d*
Log(RT)	2,500	$t(563) = -1.16$	.88	−0.05	$t(110) = -0.31$	.62	−0.03	$t(452) = -1.14$	.87	−0.05
	1,500	$t(560) = -1.28$	.90	−0.05	$t(109) = -0.66$	.75	−0.06	$t(450) = -1.09$	.86	−0.05
1/RT [a]	2,500	$t(563) = -1.31$	.91	−0.06	$t(110) = -0.51$	.69	−0.05	$t(452) = -1.21$	.89	−0.06
	1,500	$t(560) = -1.25$	.89	−0.05	$t(109) = -0.58$	.72	−0.06	$t(450) = -1.10$	.86	−0.05

[a] The *t*-test and *d*-statistics for the inverse transformation were reversed back so that the psi hypothesis would have the mean be *greater than* 0.

because their mean accuracy was not above 75%. We applied the same data transformations as in Experiment 1 to approximate normal distributions.

Confirmatory Analyses

To reiterate, the retroactive priming hypothesis states that participants have greater RT to incongruent than congruent trials, which we primarily tested with one-sample *t*-tests on difference scores. Contrary to our prediction, there was no significant retroactive priming effect for any of the four outcome variables (log and inverse transformation with 2,500 or 1,500 ms cutoff). As shown in Table 3, *p*-values were .88 or greater for the *t*-tests on the total sample. Unlike in Experiment 1, there was no significant retroactive priming effect observed in the English-language sample (*p*s > .62).

Exploratory Analyses

The second and third hypotheses state that experimenters and participants with greater experiences and beliefs would obtain greater

TABLE 4
**Results from 4 Mixed Models Analyzing Whether the Four Retroactive Priming
Outcomes Depend on Experiment Psi Belief/Experience (Hypothesis 2), Participant
Psi Belief/Experience (Hypothesis 3), or Their Interaction (Hypothesis 4)**

| | Log Data | | | | Inverse Data | | | |
| | 2,500 | | 1,500 | | 2,500 | | 1,500 | |
Predictor	F	p	F	P	F	p	F	p
Experimenter (E)	0.63	.54	3.02	.05*	1.50	.22	3.47	.03*
Participant (P)	0.79	.46	0.12	.89	0.39	.68	0.10	.91
E × P	0.67	.61	0.87	.48	0.97	.42	0.99	.41

The four models included random intercepts and fixed slopes, and maximum likelihood estimation method. Degrees of freedom (*df*) were calculated with the Satterthwaite method. Numerator df was 2 for main factors and 4 for the interaction. Denominator *df* was 525 for 2,500 ms cutoffs and 522 for 1,500 ms cutoffs. * $p < .05$

retroactive priming effects, whereas the fourth hypothesis states that there is an interaction between experimenters and participants. The two group factors were used as predictors in four mixed models (one for each of the four outcomes; log and inverse data with 2,500 ms or 1,500 ms cutoffs). Across all four outcomes, there were two significant main effects of experimenters' belief/experience, but these effects would not remain significant after a Bonferroni correction for four analyses. Nevertheless, experimenters with medium belief/experience had greater scores than those with low belief/experience ($p = .04$ for inverse and $p = .06$ for log outcome after Bonferroni correction). The overall results are shown in Table 4. In summary, we did not find clear support for a relation between self-reported belief and experience of ESP and behavioral outcome.

We then tested the hypothesis that the retroactive priming effect would be greater in sessions with a pro-psi than with an anti-psi introduction. For the log-transformed data with a 2,500 ms cutoff, there was a significant difference, $t(562) = 1.68$, $p = .05$, $d = 0.14$. Specifically, those 293 individuals who read the pro-psi statement had a mean difference score of 0.002 ($SD = 0.082$), whereas those 271 individuals who read the anti-psi statement had a mean difference score of –0.010

($SD = 0.088$). This statement effect was also significant for the 1,500 ms cutoff (log transform), $t(559) = 1.78$, $p = .04$, $d = 0.15$. In contrast, this effect was not significant for the inverse transformations, $t(562) = 1.24$, $p = .11$, $d = 0.10$ for the 2,500 ms cutoff, and $t(559) = 1.23$, $p = .11$, $d = 0.10$ for the 1,500 ms cutoff. Thus, there was mixed support for an effect of pro- vs. anti-psi statement on retroactive priming outcome according to these exploratory analyses.

As there was a small positive correlation in Experiment 1 between the variable *getting bored easily* and the log and inverse outcomes with the 1,500 ms cutoff, we examined these associations in Experiment 2. The correlation observed was close to zero and not significant for log, $r(559) = .00$, $p = .50$, and inverse, $r(559) = -.01$, $p = .62$, outcomes. We also followed up on the correlation between Stimulus Seeking and the retroactive priming effect from Experiment 1 (log transformed, 1,500 ms cutoff). The correlation between these two variables in Experiment 2 was not significant, $r(559) = .05$, $p = .14$. Thus, we could not replicate the association between getting bored easily (and Stimulus Seeking, more broadly) and the retroactive priming effect observed in Experiment 1.

DISCUSSION

Overall, the two studies did not replicate the original Bem findings for time-reverse priming on RT. Both failed to reach significance in the preplanned confirmatory hypotheses. Exploratory analyses indicated significant effects in Study 1 for the English-only condition, which is consistent with Bem's initial work. Study 2 was successful in producing a greater effect on time-reversed RT for those who received a pro-psi prompt as compared with the negative psi prompt (although this effect may be more driven by the anti-psi than the pro-psi statement). These results indicate that the brief comments of the two genuine-but-disagreeing experts in this experiment held greater sway over the participants' psi performance than did their own initial beliefs and experiences. Study 2 did not yield significant psi effect in the English-language sample.

These studies build upon previous research by exploring whether the observations about beliefs in psi may play a role in the replication of anomalous results under controlled conditions. One limitation of

these two studies is that expectancies and beliefs were evaluated using self-report questionnaires. In a future study, the role of unconscious beliefs will be assessed to further understand the role of beliefs in psi performance. The implicit association test originally developed by Greenwald et al. (1998) has shown that overt responses of participants do not necessarily reflect their unconscious beliefs. This will be the approach in Study 3 of the series (Schlitz & Delorme, 2021).

At a meta-level, the studies strongly support the feasibility of a multi-laboratory collaboration involving researchers representing different worldviews and beliefs about psi phenomena. With the aid of technology and the spirit of goodwill, these studies speak to mutual support for common interests in the empirical study of psi phenomena.

As we aim to assimilate these results, we suggest that reported findings are open to two main competing interpretations. First, initial studies reported by Bem and colleagues may have been caused by a genuine psi effect and the current experiments failed to fully replicate this finding because some aspect of the current studies disrupted the production of that effect. Although it is impossible to falsify this position, it is difficult to identify any obvious factors that might have prevented a psi effect from operating. The issue of language (and culture) reveals an important dimension; did participants who were working in other languages lack the depth of understanding about the study and the goals enjoyed by native English speakers? Perhaps the interpretations and meta-cultural dimensions of the experimental exchanges were unexpected variables. It is also possible that a more subtle, unanticipated, and uncontrolled factor may have disrupted the production of an overall effect on the main pre-registered hypothesis. For example, the study took place in diverse settings with no consistent environment, set, or setting across sub-experiments. The background and experiences of the experimenters were uncontrolled, with the exception of the interventions. A much larger study would be needed to find statistical significance across experimenters. Future studies might aim to select participants and experimenters who have shown talent at performing this task and to find ways to increase statistical power.

Second, it is possible that the results from earlier studies represented chance findings or undetected subtle artifacts and that the results obtained in the present studies accurately reflect the absence of

a psi effect based on the preplanned analyses. This is consistent with the null results reported in another recent multi-laboratory, pre-registered replication attempt with large N (Maier et al., 2020), which also tested retroactive influence (but without informing participants prior to testing that they would be tested for ESP). Our results are also consistent with the broader observation of finding smaller psychological effects in pre-registered replication attempts than in retrospective meta-analytic estimates (Kvarven et al., 2020). On the other hand, there may be hidden moderators that influence the outcomes of these replication attempts (c.f., pro- vs. anti-psi priming). It is certainly the case that the methodology employed in the current studies was more ambitious than the original studies in scope, and sophisticated in terms of the use of preregistration. This may be driven by the development of a field of study more than the techniques and procedures used in previous work—for example, rather than being driven by any concern that the previous findings were the result of any obvious artifacts. The results of Experiment 2 also speak to the importance of process-oriented work that is not driven by proof of concept.

Further, these studies provide rich fodder for sociological investigations of replication in science. This series of experiments demonstrates that it is possible to conduct fruitful collaborative research involving both skeptics and proponents, and it offers the potential of a more productive route than more traditional forms of skeptic–proponent debate (e.g., Honorton, 1985; Hyman, 1985; Schlitz et al., 2006). The collaborative project described here reduces the likelihood of perpetuating nonconstructive rhetoric because skeptics and proponents are actively engaged in the same study, and the procedures employed should minimize methodological flaws and maximize the procedures that proponents believe to be conducive to psi functioning. In addition, opportunities for explaining away the results post hoc are limited since the experiments made use of preregistered protocols. The interpretation of the data remains in the eyes of the beholder.

There are, however, several barriers that may hinder this type of collaborative venture. In many controversial areas of psychology, communities of researchers with opposing views tend not to attend the same conferences, publish in the same journals, or even read the same type of academic articles and books (Blackmore, 1989). Additional

barriers include an inherent distrust of one another fueled by ideological differences, personal beliefs, and past involvement in acrimonious debates. Our experience suggests that there is considerable value in trying to overcome these barriers and carry out systematic and collaborative ventures. It is hoped that the studies described here will encourage researchers working in other controversial areas (e.g., the role of "trance" in hypnosis, false memory syndrome, unorthodox forms of psychotherapy, and complementary and alternative medicine) to engage in similar joint projects and that such work will help advance our understanding of the phenomena underlying these controversies. Advancing such collegial endeavors in the pursuit of truth is ultimately dependent on the degree to which researchers engage with goodwill, an open mind, and an active sense of curiosity.

ACKNOWLEDGMENTS

REFERENCES

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407–425. https://doi.org/10.1063/1.3663724

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115*(1), 4–18. https://www.alice.id.tue.nl/references/bem-honorton-1994.pdf

Bem, D. J., Tressoldi, P. E., Rabeyron, T., & Duggan, M. (2016). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of

random future events. *F1000Research 4*, 1188. https://doi.org/10.12688/f1000research.7177.2

Blackmore, S. (1989). What do we really think? A survey of parapsychologists and sceptics. *Journal of the Society for Psychical Research, 55,* 251–262.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663–677. https://doi.org/10.1037/amp0000236

Collins, H. M., & Pinch, T. J. (1979). The construction of the paranormal: Nothing unscientific is happening. *The Sociological Review*, 27(1_suppl), 237–270.

Costa, P. T., Jr., & McCrae, R. R. (1992). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment: Vol. 2. Personality Measurement and Testing* (pp. 179–198). SAGE. https://doi.org/10.4135/9781849200479.n9

Eysenck, H. J. (1966). Personality and extra-sensory perception. *Journal of the Society for Psychical Research, 44,* 55–71.

Farley, F., & Farley, S. V. (1967). Extroversion and stimulus-seeking motivation. *Journal of Consulting and Clinical Psychology, 31*(2)**,** 215–216. https://doi.org/10.1037/h0024418

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.

Honorton, C. (1978). *Replicability, experimenter influence, and parapsychology: An empirical context for the study of mind.* Presented at AAAS conference, Washington, D. C.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology*, 49(1), 51–91.

Honorton, C., Ramsay, M., & Cabibbo, C. (1975). Experimenter effects in extra-sensory perception. *Journal of the American Society for Psychical Research*, 69(2), 135–149. https://psycnet.apa.org/record/1975-26533-001

Honorton, C., Ferrari D. C., & Bem, D. J. (1992). Extraversion and ESP performance: Meta-analysis and a new confirmation. In L. A. Henkel & G. R. Schmeidler (Eds.), *Research in parapsychology 1990* (pp. 35–38). Scarecrow Press.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *The Journal of Parapsychology*, 49(1), 3–49.

Johnson, M., Cronquist, A., Danielsson, B. I., & Mondejar, A. S. (1972). A test of clairvoyance especially designed for children: A study on the effect of the experimenter–participant relationship on ESP performance. *Research Letter of the Parapsychological Division of the Psychological Laboratory of the University of Utrecht*, 1–8.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology, 40*(1), 1–33. https://jeksite.org/psi/jp76.htm

Krippner, S. (1978). Commentary on "Interpersonal expectancy effects: The first 345 studies" by R. Rosenthal and D. B. Rubin. *The Behavioral and Brain Sciences, 1*(3), 377–386.

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*(4), 423–434.

Lang, P. J., Greenwald, M. K., & Bradley, M. M. (1993). *International Affective Picture System standardization procedure and results for affective judgments.* University of Florida Center for Research in Psychophysiology.

Maier, M. A., Buechner, V. L., Dechamps, M. C., Pflitsch, M., Kurzrock, W., Tressoldi, P., Rabeyron, T., Cardeña, E., Marcusson-Clavertz, D., & Martsinkovskaja, T. (2020). A preregistered multi-lab replication of Maier et al. (2014, Exp. 4) testing retroactive avoidance. *PLoS ONE 15*(8), e0238373. https://doi.org/10.1371/journal.pone.0238373

Nash, C. B. (1968). Comparison of ESP run-score averages of groups liked and disliked by the experimenter. *Journal of the American Society for Psychical Research, 62,* 411–417.

Palmer, J., & Millar, B. (2015). Experimenter effects in parapsychological research. In E. Cardeña, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 293–300). McFarland.

Pratt, J. G., Smith, B. M., Rhine, J. B., Stuart, C. E., & Greenwood, J. A. (1940). *Extrasensory perception after sixty years: A critical appraisal of the research in extra-sensory perception.* Henry Holt. https://doi.org/10.1037/13598-000

Rivers, O. B. (1950). An exploratory study of the mental health and intelligence of ESP subjects. *Journal of Parapsychology, 14*(4), 267–277. https://search.proquest.com/openview/1b0a1f4c0cab8fb3019b99c00d6ffc8b/1?pq-origsite=gscholar&cbl=1818062

Roe, C. A., Davey, R., & Stevens, P. (2006). Experimenter effects in laboratory tests of ESP and PK using a common protocol. *Journal of Scientific Exploration, 20*(2), 239–253.

Rosenthal, R. (1978, February). *Replicability and experimenter influence: Experimenter effects in behavioral research* [Paper presentation]. AAAS conference, Washington, D. C.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551–566.

Schlitz, M., & Braud W. (1997). Distant intentionality and healing: Assessing the evidence. *Alternative Therapies in Health and Medicine, 3*(6), 62–73. https://pubmed.ncbi.nlm.nih.gov/9375431/

Schlitz, M., & Delorme, A. (2021, January 6). Examining implicit beliefs in a replication attempt of a time-reversed task. *F1000 Research, 10*(5). https://doi.org/10.12688/f1000research.27169.1

Schlitz, M., Wiseman, R., Watt, C., & Radin, D. (2006). Of two minds: Sceptic-proponent collaboration within parapsychology. *British Journal of Parapsychology, 97*(3), 313–322. https://doi.org/10.1348/000712605X80704

Thouless, R. H. (1976). The effect of the experimenter's attitude on experimental results in parapsychology. *Journal of the Society for Psychical Research. 48*(767), 261–266. https://psycnet.apa.org/record/1977-11704-001

Wagenmakers, E.–J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432.

Watt, C., Schlitz, M., Wiseman, R., & Radin, D. (2005, August 11–15). Experimenter differences in a remote staring study. In *Proceedings of the 48th Annual Convention of the Parapsychological Association* (pp. 256–260).

West, D. J., & Fisk, G. W. (1953). A dual experiment with clock cards. *Journal of the Society for Psychical Research, 37*, 185–189.

White, R. A. (1977). The influence of the experimenter motivation, attitudes, and methods of handling subjects in psi test results. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 273–304). Van Nostrand Reinhold.

Zuckerman, M. (1974). The sensation seeking motive. In B. A. Maher (Ed)., *Progress in experimental personality research* (Vol. 7, pp. 79–148). Academic Press.

APPENDIX
SCRIPT FOR EXPERIMENTERS

Before the Participant Arrives

Turn on the computer, monitor, and speakers. Set the floor lamp to its lowest setting, turn off the overhead fluorescents, and put the "Experiment in Progress" sign on the laboratory door.

Start the program, fill out the opening screen, and leave the participant's opening screen up for the participant.

Put the date, time, the participant's name, and your own name on the Session Record Form. [A copy of this form will be found in the Supplementary Files folder.]

When the Participant Arrives

You and the participant should be seated in two comfortable chairs facing one another. Chat long enough to relax him or her. Don't rush. This is a very short experiment and there is plenty of time. You can tell the participant this, too, if he or she seems tense or rushed.

As soon as it feels right, explain the experimental procedure. You can paraphrase the following:

This is an experiment that tests for ESP (Extrasensory Perception). The experiment is run completely by a computer and takes about 15 minutes. [**Note: Most participants take less than 10 minutes to complete the experiment once it is under way.**]

First, you will be asked to answer a few questions about yourself. Then, on each trial of the experiment you will be shown a picture on the screen and asked to indicate as quickly as you can whether it is pleasant or unpleasant. A word will then flash on the screen very briefly. There will be 40 trials in all.

At the end of the session, I will explain to you how this procedure tests for ESP.

At this point, you can answer any questions they have. If they express any doubts about having ESP or worry that they won't do well, reassure them that we are primarily interested in testing the experimental procedure, not their own individual ESP ability. Participants should not feel pressured to perform, but neither should they feel that we are just playing around.

If your institution requires participants to sign a consent form to be in an experiment, now is the time to have them sign it. If they have the option of being paid money or receiving credit for participation in a course, now is the time to confirm that information and to record it on the Session Record Form (which will be found in the Supplementary Files Folder).

Seat the participant in the computer chair and, if necessary, help adjust the chair and the tilt of the computer screen to a comfortable position. Remember to ask if he or she prefers to have the mouse positioned to the left of the keyboard. **Check to make sure that cell phones—both yours and the participant's—are turned off.**

Show them the two keys on the keyboard with the frowning and smiling faces. Tell them that they will be using these to enter their responses. Show them how to rest the heels of their hands on the desk so that they can reach the two keys quickly and easily with their two index fingers.

Explain that the instructions will be shown on the screen but that you will be within hearing range if they have any questions or difficulties. Then move out of the experimental space.

During the Experiment

Fill out the questions in the middle section of the Session Form. These ask about the participant's demeanor and will be entered by hand into the database later. It is important that you respond to these questions now, before you know how well the participant did.

After the Experiment

The program provides feedback to participants on the final screen by informing them whether they responded more quickly or more slowly on congruent trials than on incongruent. Glance at the final screen so you can interpret their performance for them in your post-experiment debriefing.

Depending on how much detail feels appropriate for this participant, you can include some or all of the following points

— This experiment is designed to test for precognition, a form of ESP in which a person can anticipate the future.

— This experiment is a modified version of what is known in cognitive/social psychology as a **priming experiment**. In a typical priming experiment, the participant is asked to judge as quickly as possible on each trial whether a picture is pleasant or unpleasant, and the time it takes for him or her to make a response is measured. Just before each picture appears, a word is flashed very quickly on the screen. This word is called a prime. On some trials the word and the picture are matched; that is, they are either both pleasant or both unpleasant. On other trials they are mismatched. For example, a pleasant word like "beautiful" might be flashed just before a picture of a snarling dog appears on the screen. Typically people respond more quickly when the word and the picture are matched than when they are mismatched.

— In the modified version of the experiment that you just participated in, the sequence was reversed so that the word was flashed AFTER you had already made your response to the picture. This is how we tested for ESP. If people can be affected by the immediate future, then the priming word could affect their response time even though it occurs after they make their judgment about the picture. Accordingly, the ESP hypothesis is that people will respond more quickly when the priming word matches the picture than when it doesn't—even though the word has not yet been flashed. Your results showed that you did [in fact, show this ESP result] [not show this pattern, however]. We are finding that some participants show the effect and others do not, and we are attempting to discover what might produce such an effect as it is actually found in the experiment.

Thank them for their participation and reassure them that they have done just fine and given us what we needed. If they were getting paid money, pay them and have them sign the receipt form. If they express any complaints or reservations about the experiment, give them a contact number where they

can register their complaint or give them a blank copy of the consent form with a contact listed on it so they can follow up.

After the participant leaves, add any comments or observations you have about the session that might help interpret the results (e.g., participant was rushed and unfriendly, participant was about to take a final exam after the session, participant expressed suspicions that the experiment was not really about ESP). Terminate the program by pressing 'q' on the keyboard.