AMERICAN PSYCHOLOGICAL ASSOCIATION
SOCIETY OF PSYCHOLOGICAL HYPNOSIS

Psychology of Consciousness: Theory, Research, and Practice

© 2022 American Psychological Association
ISSN: 2326-5523
https://doi.org/10.1037/cns0000342

The Future Failed: No Evidence for Precognition in a Large Scale Replication Attempt of Bem (2011)

Michèle C. Muhmenthaler, Mirela Dubravac, and Beat Meier
Institute of Psychology, University of Bern

Precognition describes the ability to anticipate information about a future event before this event occurs. The goal of our study was to test the occurrence of precognition by trying to replicate three experiments of the most central study in the field (Bem, 2011, *Journal of Personality and Social Psychology*). In this study, Bem time-reversed well-established psychological effects so that a "causal" stimulus appeared after the participants gave their response. We conducted two priming experiments and a free recall experiment in the backward "precognition" version and, as a control manipulation, in the classic forward version. More than 2000 participants participated via the Internet; thus, our study had high statistical power. The results showed no precognition effects at all. We further conducted exploratory post hoc analyses on different variables and questionnaire items and found some significant effects. Further studies should validate these potentially interesting findings by using theory-driven hypotheses, preregistrations, and confirmatory data analyses.

Keywords: parapsychology, psi phenomena, confirmatory experiments, replicability, retrocausation

Supplemental materials: https://doi.org/10.1037/cns0000342.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Have you ever thought of a person you have not spoken to for years and then this person called you out of the blue? Almost everyone can report the occurrence of such an event. Some people even report experiencing this regularly, often accompanied by a strong feeling of knowing (Cleary & Claxton, 2018). Many people believe that this phenomenon is real, based on the premise that a strong connection between people can somehow become physical and therefore

Michèle C. Muhmenthaler https://orcid.org/0000-0002-7376-1164

The authors thank Andrew Clement for proofreading. No funding was received. The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

This study was approved by the Ethics Committee of the Faculty of Human Sciences at the University of Bern and followed the guidelines of the Declaration of Helsinki. Procedures were executed in conformity with the approved guidelines. People who agreed to participate were invited by email. They were informed about the purpose of the study and that their data will be treated confidentially. They were further informed that they can cancel their participation at any time. A detailed information sheet was attached to the email. Participants gave consent by clicking the first link. The study and the consent procedure were approved by the local ethical committee.

The experiment reported in this article was not formally preregistered. The data and code for this study are available at https://doi.org/10.48620/40.

The research data has the following identifier: https://doi.org/10.48620/40 (generated 22 February 2022). This repository contains the following underlying data: Experiment 1.csv (Dataset of Experiment 1), Experiment 2.csv (Dataset of Experiment 2), Experiment 3.csv (Dataset of Experiment 3).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0). We ensure that all the datasets have been deidentified in accordance with the Safe Harbor method before uploading.

Michèle C. Muhmenthaler, Mirela Dubravac and Beat Meier designed the experiments. Michèle C. Muhmenthaler and Mirela Dubravac analyzed the data. Michèle C. Muhmenthaler and Beat Meier wrote the article and all authors approved the final article for submission.

Correspondence concerning this article should be addressed to Michèle C. Muhmenthaler, Institute of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland. Email: michele.muhmenthaler@unibe.ch

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

perceptible. According to this view, everything relies on energy, even thoughts, and thus, it is conceivable that they can be sent and received, especially by pairs of strongly connected people (Erickson, 2011; Radin et al., 2017). However, as this phenomenon cannot be explained in terms of any known physical, psychological, or biological mechanism, it counts as one of several *psi* phenomena. *Psi* phenomena are defined as anomalous processes of energy or information transfer which conflict with fundamental principles of physics (Bem & Honorton, 1994; Cardeña, 2018). Although *psi* experiences have always been a part of human history, they have been exposed to skepticism since the advent of science (Franklin et al., 2014). In the present study, we tried to replicate three experiments of the probably most discussed study in the field of *psi* phenomena (Bem, 2011) in a course which enabled us to gather data from more than 2000 participants. The goal of our study was to use this large sample from the general population to test the occurrence of *precognition,* a *psi* phenomenon which denotes a conscious awareness of the future (Honorton & Ferrari, 1989).

Precognition describes the ability to anticipate information about a future event before this event actually occurs (Bem & Honorton, 1994). Thus, the existence of precognition would imply that the time axis can be reversed. This finding would challenge the second law of thermodynamics, which states that the arrow of time can never be reversed (Franklin et al., 2014; Schwarzkopf, 2014; Sheehan, 2006). Although seemingly impossible, Bem (2011) provided empirical evidence supporting precognition in nine experiments which were conducted under well-controlled laboratory conditions. In these experiments, Bem reversed the causal direction of several well-established psychological effects; thus, the participants gave their responses *before* a "causal" stimulus occurred.

In two experiments (Bem, 2011; Experiments 3 and 4), Bem time-reversed an affective priming procedure. In a typical priming experiment, participants have to judge a stimulus according to different criteria as fast as possible (Klauer & Musch, 2003). Shortly before the stimulus appears, a prime (which can be a word, a picture, etc.) is briefly flashed (Hermans et al., 2010). Priming takes place when the prime and the target are semantically related (i.e., congruent), which leads to faster processing of the target due to pre-activation of the semantic network. In the (backward) precognition version, the prime is presented *after* the participant judged the target stimulus. In Bem's Experiment 3, the participants had to judge pictures as positive or negative. Afterward, one of the two prime words, "beautiful" or "ugly," was briefly flashed. In Bem's Experiment 4, a semantically related fixed prime word pair (one positive, one negative) was assigned to each picture. For example, a picture of a plane crash was paired with the positive prime word *gorgeous* and the negative prime word *deadly*. The *psi* hypothesis was that responses for congruent trials should be faster than for incongruent trials due to a "retro-activation" of the semantic network from the future into the past.

In two further experiments (Bem, 2011; Experiments 8 and 9), Bem time-reversed a facilitation of free recall procedure. A typical facilitation of free recall experiment investigates whether rehearsing a subset of previously encountered words makes them easier to recall than unrehearsed words (Brown, 1968). In Bem's study, the participants learned 48 words by visualizing them for 3 s. Then they were asked to recall as many of these words as possible. After the recall test, the participants rehearsed a randomly chosen subset of the initially learned words. The *psi* hypothesis was that rehearsing the subset of words *after* a free recall test would improve free recall performance due to retroactive facilitation (Bem, 2011).

Although reporting empirical evidence for precognition in eight of nine time-reversed experiments, Bem did not provide a reasonable theoretical explanation for the effects. He emphasized that he used the term precognition in a purely descriptive manner, and that explaining underlying mechanisms or theories was beyond the scope of his study (Bem, 2011; Bem & Honorton, 1994). However, as good science should be both theory-driven and evidence-based, a lack of a reasonable theoretical framework is problematic (McMullin, 2008). Although different approaches such as evolutionary advantages (Cohn, 1999; Savva, 2014), intuition (Bernstein, 2005; Erickson, 2011), entropic considerations (Marwaha & May, 2015), and quantum mechanisms (Hameroff, 2012; Hameroff & Penrose, 2014; Hawking & Penrose, 2010; Maier et al., 2014) have been proposed, there exists no theory that can provide a satisfactory explanation for precognition.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

It is therefore astonishing that some research groups have found empirical evidence for the existence of precognition. The accumulated evidence includes successful replication attempts of Bem's study, including significant meta-analyses of forced-choice guessing and precognition experiments and presentiment experiments indicating that the human body can detect stimuli occurring in the future (Bem et al., 2015; Honorton & Ferrari, 1989; Mossbridge et al., 2012, 2014; Radin, 2004; Radin et al., 2017). The authors of one meta-analysis of 90 replication attempts of Bem's experiments from 33 laboratories in 14 countries reported an overall effect of $d = 0.18$ and a Bayes factor exceeding the criterion value of 100 for decisive evidence in support of the *psi* hypothesis (Bem et al., 2015).

However, many studies have also reported failed replication attempts of Bem's experiments (Barušs & Rabier, 2014; Galak & Nelson, 2010; Jolij & Bierman, 2019; Rabeyron, 2014), sparking a debate about the "replication crisis" (Rabeyron, 2020; Wittmann et al., 2021). For example, in a meta-analysis on retroactive free recall experiments by Galak et al. (2012), including the results of seven of their own experiments, the two experiments by Bem (2011), and the results of ten experiments conducted by other researchers, an overall effect nonsignificantly different from zero ($d = .04$) was found; thus, the null hypothesis could not be rejected. Moreover, several authors criticized methodological issues such as a bias in favor of the hypothesis, selective data collection, too liberal statistical analyses, or other inappropriate research practices associated with precognition and presentiment studies (Gauvrit, 2011; Houran et al., 2018; Hyman, 1985; LeBel & Peters, 2011; Wagenmakers et al., 2011).

The Present Study

The present study was conducted in an undergraduate class at the University of Bern. Due to the straightforward experimental designs and due to didactic reasons, we chose the priming experiments and a free recall experiment from Bem (2011) for replication. The study was conducted online due to the COVID-19 pandemic in 2020. As we were able to gather data from more than 2000 participants, the statistical power of the study was large. This is important to detect a small deviation from the null hypothesis, should such a deviation exist (Hallahan & Rosenthal, 1996). In order to calculate power, we used the smallest effects that would be plausible and still interesting, and the results revealed that our study was suited to find such small potential effects (cf. Dienes, 2021; Gelman & Carlin, 2014; Perugini et al., 2014). In our study, the participants conducted two of the three experiments in the backward "precognition" version and one of the experiments in the classic forward version. After completing the three experiments, the participants responded to a questionnaire which included demographic items, several scales, and questions. We included a short version of the Big Five questionnaire in order to find potential relationships between personality traits and *psi* performance (Muck et al., 2007). For example, *psi* performance has been found to be correlated with emotional stability and aspects of conscientiousness and openness (Cardeña et al., 2015; Zingrone et al., 1999). We also included questions about stimulus seeking, a component of extraversion, as Bem found significant correlations between this variable and *psi* performance (Bem, 2011). Moreover, we added questions about magical and unusual beliefs (Klein et al., 1997; Rattet & Bursik, 2001) and questions about impulsivity. Finally, we assessed whether the participants were under the influence of psychoactive substances at time of testing and whether they regularly consumed psychoactive substances, as experiences of precognition have been reported after the use of psychoactive drugs (Luke, 2006).

Our experimental design involved a large number of additional exploratory analyses. In traditional statistics, there is always the possibility to commit a Type I error, that is, the rejection of the null hypothesis although it is true (Voelkl, 2019). The probability of committing such an error is equal to the significance level. For example, when generating a random number between 0 and 100, we would expect that at least one number less than or equal to five should be generated after 20 tries. Following on this, a p value less than .05 indicates that in one of 20 tests, a result will be significant by chance (Morgan, 2007). Moreover, when conducting multiple tests, the probability of falsely rejecting the null hypothesis is increased and the α level should therefore be corrected, as proposed by Bonferroni (Cabin & Mitchell, 2000). We therefore corrected the α level ($p = .05/19 = .003$), as we conducted 19 post

hoc analyses on demographic variables (i.e., gender) and questionnaire items.

Method

Participants

The 2,164 (1,276 female, 874 male, 14 other) participants were recruited by undergraduate students for a research course at the University of Bern. Their age ranged from 18 to 93 years ($M = 28.90$, $SD = 13.66$). They were told that they were recruited to participate in a study about imagination; the *psi* hypothesis was not mentioned. The criteria for participation were to be German speaking, to be older than 18, and to not be a psychology student. People who agreed to participate were invited by email. They were informed about the purpose of the study and were told that their data would be treated confidentially. They were further informed that they could cancel their participation at any time. A detailed information sheet was attached to the email. Participants gave consent by clicking the first link. The study and the consent procedure were approved by the local ethics committee. We computed the power of each experiment as a function of the sample size and the effect size based on Bem ($d = 0.22$; Bem, 2011; Faul et al., 2007). With an estimated effect size of $d = 0.22$, the analyses revealed that the power of each experiment exceeded 99%. However, power analyses should also be conducted using the smallest effect sizes which were interpreted as meaningful in previous studies (Dienes, 2008, 2021). Toward this goal, we used the smallest effects which could be interpreted in favor of the *psi* hypothesis of the meta-analysis by Bem and colleagues (Albers & Lakens, 2018; Bem et al., 2015).[1] We used the lower boundary of a 90% confidence interval (CI) of these effect sizes as a threshold for a potential interesting effect. The threshold was calculated separately for each experiment as described below.

Materials

For the priming experiments, the pictures were selected from the International Affective Picture System (IAPS; Lang et al., 2005). The IAPS is a set of 820 digitized photographs that have been rated on 9-point scales for valence and arousal by both male and female raters. The selected pictures were clearly positively or negatively connoted (e.g., a beautiful beach or a car accident). The same pictures were used in the forward and in the backward version of each priming experiment. In Experiment 1, the replication of Bem's Experiment 3, we chose 36 pictures. The words "schön" and "hässlich" (German for "beautiful" and "ugly") were used as primes. In Experiment 2, the replication of Bem's Experiment 4, we used the same materials as in Bem's original study. The prime words were selected to be semantically relevant to the picture; for example, a picture of a menacing pit bull was paired with the positive prime of "friendly" and the negative prime of "threatening". We used the same prime words as Bem, translated into German (Wittmann et al., 2021).

For the free recall experiment (Experiment 9 of Bem's study), the materials consisted of 48 common nouns of the four categories clothes, animals, musical instruments, and food, with 12 words per category. The nouns consisted of 4–10 letters. The words were printed in a black sans-serif 32-point and bold font.

After the experiments, participants completed a questionnaire. It included the 10-item personality inventory, which is a short version of the Big Five personality traits translated into German (Muck et al., 2007), and two questions about stimulus seeking developed by Bem (2011). Responses were recorded on 7-point scales that ranged from "absolutely untrue" to "absolutely true" and were averaged into a single score for stimulus seeking ranging from 1 to 7. The questionnaire included the scale "magical and unusual beliefs" from the Schizotypal Personality Questionnaire (Klein et al., 1997; Raine & Raine, 1991) with seven items. Responses were averaged into a single score between 1 and 2 (reverse scored). "Impulsivity" was assessed with the two statements "I am impulsive" and "I easily get carried away by an exciting moment." Responses were recorded on a 3-point scale and were averaged into a single score for impulsivity. Moreover, we added four questions about the consumption of psychoactive substances. We asked the participants whether they were influenced by psychoactive substances at the

[1] Dataset citation: Bem, D., Tressoldi, P. E., Rabeyron, T. and Duggan, M. (2016). Dataset 1 in: Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research*, 4, Article 1188 (https://doi.org/10.5256/f1000research.7177.d105136).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

moment and if so, which substances. Finally, we asked whether the participants consumed psychoactive substances regularly and which substances they had regularly consumed in recent months.

General Procedure

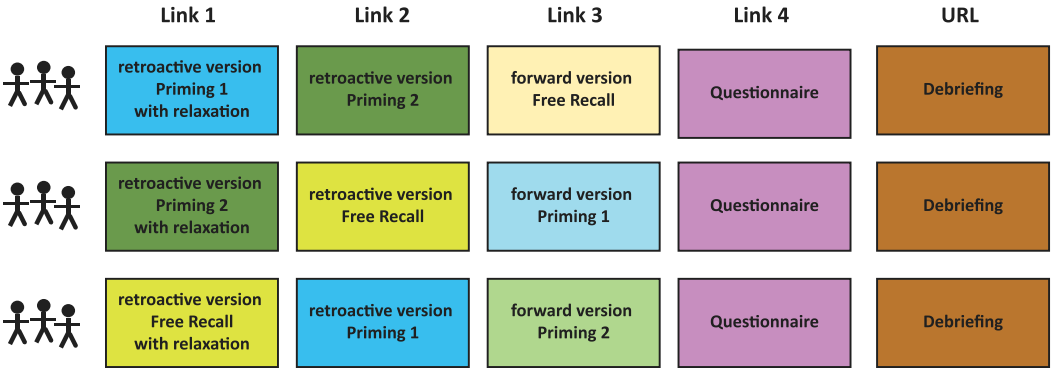
The participants received four web links per email in which they were instructed to conduct the experiments on a desktop computer or laptop in a quiet environment. The web links could be opened in any web browser. The first link contained one of the three experiments in the backward version, and it included a 3-min relaxation phase. During this relaxation period, the screen displayed colored galaxies and stars while peaceful music played (cf. Bem, 2011). The second link contained one of the two remaining experiments in the backward version, without the relaxation phase. The third link contained the remaining experiment in the classic forward version. The fourth link contained the questionnaire. After completing the questionnaire, the participants were redirected to a debriefing site. Thus, all the participants conducted both priming experiments and the free recall experiment in one of three possible orders (see Figure 1). This configuration was chosen for optimal use by the participants. The data of the two links of the backward versions of each experiment were collapsed (e.g., "Priming 1 with relaxation" and "Priming 1"), as these experiments were identical with the exception of the presence or absence of the relaxation phase.

Analyses

In order to investigate "*psi* performance," we conducted one-tailed paired samples *t* tests with the experimental conditions (congruent vs. incongruent in the priming experiments; practiced vs. control word in the free recall experiment) on the dependent variables, response time (RT) and number of recalled words, respectively. We conducted two exploratory 2 × 2 analyses of variances (ANOVAs) to investigate whether *psi* performance interacted with gender (as in Bem's study) or regular consumption of psychoactive substances. For the correlations, we computed "*psi* factors" which consisted of the average RT incongruent minus RT congruent in the priming experiments and the average difference score for practiced and control words in the free recall experiment. Using the *psi* factors, we computed correlations with the items of the personality inventory (Muck et al., 2007) and the average scores for stimulus seeking, unusual and magical beliefs, and impulsivity.

We also specified the smallest effect that was interesting for theoretical reasons. For example, if somebody wants to lose weight, a smaller effect than half a kilo over a month may probably not be of clinical interest (Dienes, 2021). In order to specify the smallest effect of interest for the *psi* hypothesis, we used the bottom limit of the 90% CI of the smallest and still somehow interesting

Figure 1
General Procedure



Note. Data of links depicted in the same color were collapsed. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

effect size which we found for the specific studies (priming and free recall, respectively) in the meta-analyses of Bem et al. (2015). Using this smallest effect size that is plausible, we calculated power analyses for each experiment.

Experiment 1

Experiment 1 was a replication attempt of Experiment 3 by Bem (2011).

Procedure

The participants were instructed to judge pictures as pleasant or unpleasant by pressing the "A" or "L" key as fast as possible. The key assignment was counterbalanced across participants. The participants were told that directly after judging the picture, a word would briefly be flashed followed by a Hubble photograph, both requiring no action. Then, they were told that they would have to judge the next picture. The 36 pictures were shown in random order, and a prime word was randomly selected on each trial after the participant had responded to the picture. As a result, congruent and incongruent trials were randomly presented and did not necessarily occur in equal numbers (Bem, 2011).

Retroactive Version

In the retroactive version, a fixation cross was presented for 1,000 ms at the beginning of a trial. Then, a picture was presented until a response was chosen by pressing the "A" or "L" key. After a blank screen of 300 ms, one of the two prime words, "beautiful" or "ugly," was randomly chosen and flashed for 500 ms. After a blank screen of 1,000 ms, a Hubble photograph was presented for 2000 ms (see Figure 2). Then, the next trial appeared accordingly. The priming experiment involved 36 trials and lasted about 8 min.

Forward Version

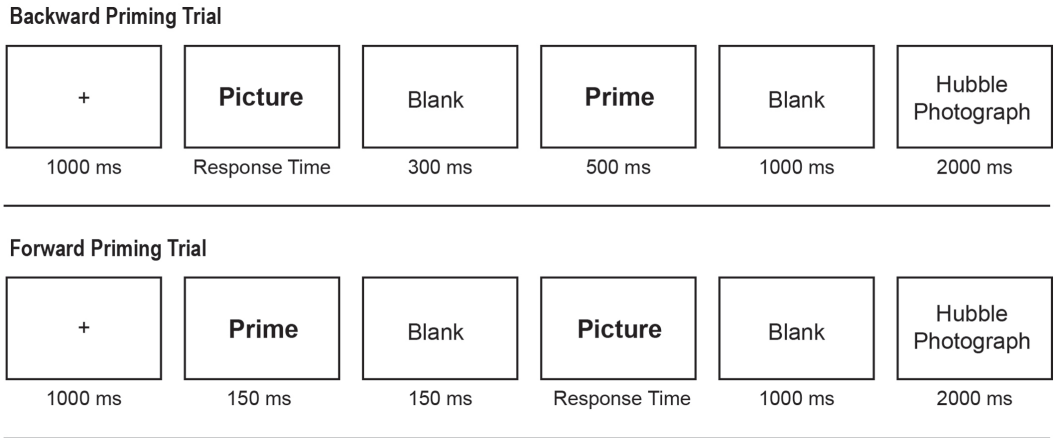
The classic forward version was identical to the backward version with the exception that the prime was flashed before the picture was presented. The participants were instructed accordingly. The prime was presented for 150 ms followed by a blank screen of 150 ms (see Figure 2).

Results

Data Preparation

In the backward version, we excluded data from six participants with an error rate >25% (cf. Bem, 2011). An error meant that a participant judged a positively connoted picture as unpleasant or a negatively connoted picture as pleasant. The final sample consisted of 727 participants (443 female, 279 male, 5 other). Their age ranged from 18 to 83 years ($M = 28.50$, $SD = 13.60$). In the classic forward version, we excluded seven participants with an error rate >25%; the final sample consisted of 705

Figure 2
Depiction of the Time Sequence of a Trial in the Backward and in the Forward Version in Experiment 1



This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

participants (423 female, 278 male, 4 other). Their age ranged from 18 to 79 ($M = 29.4$ $SD = 14.02$). The "error" trials in judging the picture to be pleasant or unpleasant were excluded (5.2% in the backward version, 4.9% in the forward version). Trials with very short (<200 ms) or very long response times (>2,500 ms) were considered as outliers and excluded from the analysis (7.5% of trials in the backward version and 4.5% of trials in the forward version). Because RT data are right skewed, the RTs were transformed using an inverse transformation (1/RT; cf. Bem, 2011).

We calculated power with respect to the smallest effect of interest that is plausible given its scientific context (Dienes, 2021; Perugini et al., 2014). Toward this goal, we chose the effect size by Rabeyron and Watt (2010; $t = 1.32$, $df = 154$, $p = .09$, $d = 0.22$).[2] Although the main effect was not significant, we could not rule out that the effect was potentially interesting. By using the bottom CI (90%) of this effect size ($d = 0.125$), an α level of .05, and a power of .90, a sample size of $N > 550$ would be required to detect an effect. With our sample of 727 participants, we had a power of even .95. Thus, our sample was large enough to find an effect in this magnitude if it was there.

For the correlations, we also calculated power with respect to a small but meaningful correlation between a variable of our questionnaire and *psi* performance. Toward this goal, we chose the correlation between extraversion and extrasensory perception from the meta-analysis by Honorton et al. (1998; $r = 0.18$, $p = .008$). By using the bottom CI (90%) of this correlation ($r = 0.12$), the corrected α level ($p = .003$) and a power of .90., a sample size of $N > 830$ would be required to detect an effect. With our sample of 727 participants, we had a power of .69.

Retroactive Priming

The one-tailed paired samples t test revealed that congruent ($M = 1{,}217$ ms, $SD = 289$ ms) and incongruent trials ($M = 1{,}215$ ms, $SD = 282$ ms) did not differ, $t(726) < 1$, $p = .512$, $d = 0.001$. Thus, we did not find evidence for precognition.

Exploratory ANOVAs. We conducted a 2×2 ANOVA with the factors gender (male vs. female; others were excluded) and congruence. There were no significant gender differences with regard to a precognition effect, $F(1, 720) < 1$, $p = .707$, $\eta_p^2 < .01$. We conducted a 2×2 ANOVA with the factors congruence and regular substance consumption (yes vs. no). Only 75 participants reported consuming psychoactive substances regularly. Substance consumption and precognition did not interact significantly, $F(1, 725) < 1$, $p = .813$, $\eta_p^2 < .01$.

Correlation Analyses. An overview of the correlations is depicted in Table 1. The correlation between the item "sympathetic, warm" of the short version of the Big Five questionnaire, and the *psi* factor was initially significant ($r = 0.075$, $p = .042$). However, the correlation did not survive α-level correction ($p > .003$).

Forward Priming

The one-tailed paired samples t test revealed that the participants responded faster to congruent ($M = 1{,}033$ ms, $SD = 274$ ms) than to incongruent trials ($M = 1{,}037$ ms, $SD = 274$ ms), $t(704) = 1.84$, $p = .033$, $d = 0.07$. Thus, the expected priming effect was found, although the effect was very small (4 ms).

Exploratory Post Hoc Analyses. The ANOVA with repeated measures with the factors congruence and gender revealed that the interaction almost reached significance, $F(1, 699) = 3.79$, $p = .052$, $\eta_p^2 < .01$. Men had slower reaction times on congruent ($M = 1{,}065$ ms, $SD = 269$ ms) than on incongruent trials ($M = 1{,}058$ ms, $SD = 269$ ms), $t(277) < 1$, $p = .747$, $d = 0.02$, whereas for women, the opposite pattern was found (congruent: $M = 1{,}010$ ms, $SD = 275$ ms; incongruent: $M = 1{,}021$ ms, $SD = 278$ ms), $t(422) = 2.62$, $p = .009$, $d = 0.127$. This indicates that only women showed a reliable priming effect. Overall, women responded faster than men, $F(1, 699) = 5.01$, $p = .025$, $\eta_p^2 < .01$. The priming effect did not interact with substance consumption, $F(1, 703) = 1.04$, $p = .309$, $\eta_p^2 < .01$.

Experiment 2

Experiment 2 was a replication of Experiment 4 by Bem (2011).

[2] Three studies with smaller positive effect sizes of the meta-analyses could not be found.

Table 1
Correlations Between the Psi Factor and Questionnaire Items

Item	Value	Priming 1	Priming 2	Free recall
Extraverted, enthusiastic	Pearson's *r*	0.050	−0.040	0.027
	P value	0.175	0.130	0.305
Critical, quarrelsome	Pearson's *r*	−0.025	0.003	−0.014
	P value	0.507	0.898	0.594
Dependable, self-disciplined	Pearson's *r*	0.003	−0.036	−0.024
	P value	0.932	0.171	0.364
Anxious, easily upset	Pearson's *r*	−0.016	−0.014	−0.061
	P value	0.675	0.596	0.023
Open to new experiences, complex	Pearson's *r*	−0.026	0.009	−0.018
	p value	0.491	0.742	0.511
Reserved, quiet	Pearson's *r*	0.011	0.002	−0.023
	p value	0.774	0.943	0.400
Sympathetic, warm	Pearson's *r*	0.075	−0.044	−0.046
	p value	0.042	0.101	0.086
Disorganized, careless	Pearson's *r*	0.027	0.016	0.005
	p value	0.460	0.560	0.850
Calm, emotionally stable	Pearson's *r*	−0.036	0.030	0.063
	p value	0.330	0.260	0.018
Conventional, uncreative	Pearson's *r*	−0.022	0.034	−0.002
	p value	0.547	0.203	0.945
Stimulus seeking	Pearson's *r*	−0.061	0.006	−0.040
	p value	0.101	0.817	0.135
Unusual and magical beliefs	Pearson's *r*	−0.027	−0.081	−0.010
	p value	0.465	**0.002**	0.712
Impulsivity scale	Pearson's *r*	0.021	−0.007	0.012
	p value	0.580	0.788	0.651

Note. Correlation factors and *p* values for all experiments. Significant results (after α-level correction) are highlighted in bold.

Procedure

Following Bem (2011), the procedure was almost identical to Experiment 1, with the following changes (cf. Bem, 2011). One fixed positive prime and one fixed negative prime were assigned to each picture prior to the experiment. These primes were selected to be semantically relevant to the picture. For example, a picture of a basket of fruit was paired with the positive prime "luscious" and the negative prime "bitter." Moreover, the duration of the fixation point was increased from 1,000 ms to 1,500 ms and the time between trials (during which the Hubble photograph appeared on the screen) was decreased from 2000 ms to 1,500 ms (see Figure 3).

Results

Data Preparation

In the backward version, we excluded 21 participants due to more than 10 error trials (>25%; cf. Bem, 2011). An error meant that a participant judged a positively connoted picture as unpleasant or a negatively connoted picture as pleasant. The final sample consisted of 1,414 participants (846 female, 559 male, 9 other). Their age ranged from 18 to 83 years ($M = 28.78$, $SD = 13.52$). In the classic forward version, we excluded 14 participants due to an error rate >25% (cf. Bem, 2011). The final sample consisted of 700 participants (400 female, 298 male, 2 other); their age ranged from 18 to 81 years ($M = 29.00$, $SD = 13.63$). In the backward version, we excluded 5.3% of error trials and 4.9% of slow trials. In the forward version, we excluded 5.0% of error trials and 5.1% of slow trials. The analyses were similar to Experiment 1.

As in Experiment 1, we chose the effect size by Rabeyron and Watt (2010; $t = 1.32$, $df = 154$, $p = .09$, $d = 0.22$) as the smallest effect of interest that is plausible (Dienes, 2021; Perugini et al., 2014). By using the bottom CI (90%) of this effect size ($d = 0.125$), an α level of .05, and a power of .90, a sample size of $N > 550$ would be required to detect an effect. With our sample of 1,414 participants, we had a power of even .99. Thus, our

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3
Depiction of the Time Sequence of a Trial in the Backward and in the Forward Version in Experiment 2

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

sample was large enough to find an effect in this magnitude if it was there.

For the correlations, as in Experiment 1, we calculated power by using the correlation between extraversion and extrasensory perception from the meta-analysis by Honorton et al. (1998; $r = 0.18$, $p = .008$). By using the bottom CI (90%) of this correlation ($r = 0.12$), the corrected α level ($p = .003$) and a power of .90., a sample size of $N > 830$ would be required to detect an effect. With our sample of 1,414 participants, we had a power of even .96. Thus, our sample was large enough to find an effect in this magnitude if it was there.

Retroactive Priming

The one-tailed paired samples *t* test revealed that congruent ($M = 1{,}170$ ms, $SD = 261$ ms) and incongruent trials ($M = 1{,}168$ ms, $SD = 262$ ms) did not differ, $t(1413) < 1, p = .227, d = 0.034$. Thus, we did not find evidence for precognition. The *psi* factor (incongruent trials minus congruent trials per participant) was on average −2 ms ($SD = 114$ ms).

Exploratory ANOVAS. We conducted a 2×2 ANOVA with the factors gender (male vs. female; others were excluded) and congruence. There were no significant gender differences with regard to a precognition effect, $F(1, 1403) = 1.07$, $p = .301$, $\eta_p^2 < .01$. The ANOVA with the factors congruence and regular substance consumption revealed that the factors did not interact significantly, $F(1, 1412) < 1$, $p = .889$, $\eta_p^2 < .01$.

Correlation Analyses. An overview is presented in Table 1. The correlations between the score for unusual and magical beliefs and the *psi* factor reached significance ($r = 0.081$, $p = .002$). The correlation survived α-level correction for multiple testing ($p < .003$). We further analyzed this finding by selecting participants who scored high in unusual and magical beliefs (i.e., endorsed all seven questions of the scale). While the descriptive data suggested that these participants responded faster to congruent ($M = 1{,}160$ ms, $SD = 239$ ms) than to incongruent trials ($M = 1{,}168$ ms, $SD = 245$ ms), the one-tailed paired samples *t* test was not significant, $t(438) < 1$, $p = .184$, $d = 0.043$.

Forward Priming

The one-tailed paired samples *t* test revealed that the participants responded faster to congruent ($M = 1{,}051$ ms, $SD = 259$ ms) than to incongruent trials ($M = 1{,}083$ ms, $SD = 255$ ms), $t(699) = 10.2$, $p < .001$, $d = 0.39$. Thus, the expected priming effect was found. The priming factor was on average 32 ms ($SD = 120$ ms).

Exploratory ANOVAS. The ANOVA on congruence and gender (male vs. female, others excluded) revealed no significant gender differences with regard to a priming effect, $F(1, 696) < 1$, $p = .433$, $\eta_p^2 < .01$. The ANOVA with the factors congruence and regular substance consumption revealed that consumption was

not related to priming performance, $F(1, 698) < 1$, $p = .265$, $\eta_p^2 < .01$.

Experiment 3

The experiment replicated the free recall experiment from Bem's study (Experiment 9).

Procedure

The participants were first shown a set of words and given a free recall test of the words. Then they practiced a randomly selected subset of these words. Our rehearsal task was one of two rehearsal tasks in Bem's Experiment 9. At the beginning, the participants were instructed that they would be shown 48 words from the four categories foods, animals, musical instruments, and clothing, and that they would have to visualize the concept of each word as it appeared on the screen. The participants were then shown the 48 nouns one at a time for 3 s each. The words were presented block-wise, and the blocks were presented in randomized order. Within a category block, the words were presented in randomized order. After 48 trials, a surprise free recall test was applied. The participants were instructed to type all the words they remembered in any order and to press the "ENTER" key when they had completed this task. The time for this task was not limited. After the participant completed the recall test, the computer randomly selected six words from each of the four categories to serve as practice words, with the remaining 24 words serving as control words. The participants were again instructed to visualize the concept of each word as it appeared on the screen, and each word was then displayed for 3 s. Again, the words were presented in four category blocks, and the blocks and the words within blocks were presented in randomized order. In the forward version, the practice phase was administered before the free recall test took place.

Results

Data Preparation

In the backward version, we tested 1,395 participants (809 female, 578 male, 8 other). Their age ranged from 18 to 93 years ($M = 28.84$, $SD = 13.30$). In the classic forward version, we tested 717 participants (433 female, 278 male, 6 other). Their age ranged from 18 to 83 years ($M = 28.64$, $SD = 13.67$). We excluded intrusions and word repetitions from analyses. For the correlations, following Bem's design, we computed a difference score which was a measure for *precognition* performance. This measure was the weighted differential recall (DR) score, defined as the number of practice words recalled minus the number of control words recalled ($P - C$) multiplied by the participant's overall recall score ($P + C$). The reason for this measure was to give more weight to participants who recalled more words (Bem, 2011). The DR score refers to the *psi* factor, and we computed the correlations with this value.

To find the smallest plausible effect size associated with a retroactive free recall study, we consulted the meta-analysis by Bem et al. (2015) again (Dienes, 2021; Perugini et al., 2014). The experiment with the smallest effect size which was significant (one-sided tested) was the one by Galak et al. (2012; Experiment 4, $t = 1.77$, $df = 108$, $p = .04$, $d = 0.17$). By using the bottom CI (90%) of this effect size ($d = 0.058$), an α level of .05, and a power of .90, a sample size of $N > 1840$ would be required to detect an effect. With our sample of 1,395 participants, we had a power of .70.

For the correlations, we calculated power with respect to a small but meaningful correlation in the literature between a variable of our questionnaire and retroactive free recall (Bem et al., 2015).[3] Toward this goal we chose the correlation between stimulus seeking and the *psi* factor in Experiment 8 (free recall) of Bem (2011; $r = 0.22$, $p = .015$). By using the bottom CI (90%) of this correlation ($r = 0.14$), the corrected α level ($p = .003$) and a power of .90, a sample size of $N > 678$ would be required to detect an effect. With our sample of 1,395 participants, we had a power of .99. Thus, our sample was large enough to find an effect in this magnitude if it was there.

Retroactive Facilitation of Free Recall

The one-tailed paired samples t test revealed that the participants recalled the same number of practiced words ($M = 11.4$ $SD = 3.9$) and control words ($M = 11.5$, $SD = 3.9$), $t(1394) = 1.08$, $p = .860$, $d = -0.03$. The average DR score was 2.48 ($SD = 77.80$). Thus, we did not find evidence for precognition.

[3] Many papers reported in the meta-analysis are unpublished, do not report correlations or have very small samples.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Exploratory ANOVAS. The ANOVA on word type (practiced vs. control) and gender (male vs. female; others were excluded) revealed no gender differences regarding a precognition effect, $F(1, 1385) < 1, p = .729, \eta_p^2 < .01$. Overall, women recalled significantly more words ($M = 23.9$, $SD = 6.5$) than men ($M = 21.4$, $SD = 7.5$), $F(1, 1385) = 44.8$, $p < .001$, $\eta_p^2 = .03$. The ANOVA with the repeated measures words (practiced vs. control) and regular substance consumption (yes vs. no) revealed a significant interaction, $F(1, 1374) = 6.60$, $p = .010$, $\eta_p^2 = .01$. Additional one-tailed paired samples t tests revealed that participants who reported consuming psychoactive substances showed a precognition effect (practiced words: 11.4, $SD = 4.0$, vs. control words: 10.9, $SD = 3.9$), $t(170) = 1.80$, $p = .037$, $d = 0.138$, whereas nonconsumers did not show a precognition effect (practiced words: 11.4, $SD = 3.9$, vs. control words: 11.6, $SD = 3.9$), $t(1205) = 1.86$, $p = .971$, $d = 0.055$. The initially significant interaction did not survive α-level correction ($p > .003$).

Correlation Analyses. An overview is presented in Table 1. Two items of the short version of the Big Five questionnaire (Muck et al., 2007) were initially correlated with the *psi* factor: The item "anxious, easily upset" was negatively correlated ($r = -0.061, p = .023$), and the item "calm, emotionally stable" ($r = 0.063$, $p = .018$) was positively correlated with *psi* performance. Both personality items were negatively intercorrelated ($r = —0.523$, $p < .001$) and reflected the factor "emotional stability" of the Big Five questionnaire (Costa, 1992; Muck et al., 2007). The correlations with the *psi* factor did not survive α-level correction ($p > .003$).

Forward Facilitation of Free Recall

The one-tailed paired samples t test revealed that the participants recalled more practiced words ($M = 14.8$, $SD = 3.3$) than control words ($M = 7.8$, $SD = 4.4$), $t(716) = 49.1$, $p < .001$, $d = 1.83$. Thus, the expected effect of facilitation of recall was found. The average DR score was 157.25 ($SD = 100.50$).

Exploratory ANOVAS. The ANOVA with repeated measures with the factors word type and gender (male vs. female, others removed) revealed a significant interaction, $F(1, 709) = 5.22$, $p = .023$, $\eta_p^2 = .01$. Post hoc tests revealed that the interaction was significant due to the difference in practiced words (women: $M = 15.3$, $SD = 4.4$; men: $M = 14.0$, $SD = 4.5$), $t(709) = 4.00$, $p < .001$, $d = 0.31$. Performance for control words did not differ significantly, although women again recalled more of these words than men (women: $M = 8.1$, $SD = 4.3$, men: $M = 7.4$, $SD = 4.4$), $t(709) = 1.95$, $p = .052$, $d = 0.15$. Overall, women recalled significantly more words (23.5, $SD = 7.7$) than men (21.5, $SD = 8.0$), $F(1, 709) = 10.9$, $p < .001$, $\eta_p^2 = .02$.

We conducted an ANOVA with repeated measures with the factors word type (practiced vs. control) and substance consumption (yes vs. no). The interaction was significant, $F(1, 715) = 5.50$, $p = .019, \eta_p^2 = .01$. Post hoc analyses revealed that participants who reported consuming psychoactive substances showed a smaller facilitation effect (practiced words: 14.5, $SD = 3.9$, vs. control words 8.4, $SD = 3.9$), $t(74) = 12.00$, $p < .001$, $d = 1.38$) than nonconsumers (practiced words: 14.9, $SD = 3.9$, vs. control words 7.8, $SD = 4.0$), $t(629) = 47.72$, $p < .001$, $d = 1.90$).

Discussion

Precognition is one of several *psi* phenomena. It denotes conscious awareness of events that happen in the future. Precognition conflicts with our fundamental principles of causality and its existence would overturn basic physical and biological concepts (Bem & Honorton, 1994). The goal of this study was to replicate three experiments of Bem's controversial work in a large-scale study (Bem, 2011). We were able to gather data from more than 2000 participants participating via the Internet due to the COVID-19 pandemic in 2020. Conducting the study online has the advantage that potential experimenter effects are eliminated. Indeed, the role of the experimenter in achieving successful *psi* results has been debated in the research community (Kennedy & Taddonio, 1976).

In the backward priming experiments, performance on congruent and incongruent trials did not differ. The classic forward versions revealed a small (Experiment 1) and a robust priming effect (Experiment 2). The robust priming effect in Experiment 2 supported the decision to reject the precognition hypothesis, as a priming effect seemed to be a precondition for any potential precognition effects. In the free recall experiment,

the participants recalled the same number of control and practiced words. The classic forward version showed that the task manipulation was effective; practiced words were more often recalled than control words. Together, our confirmatory data analyses revealed no evidence for precognition at all (see Figure 4).

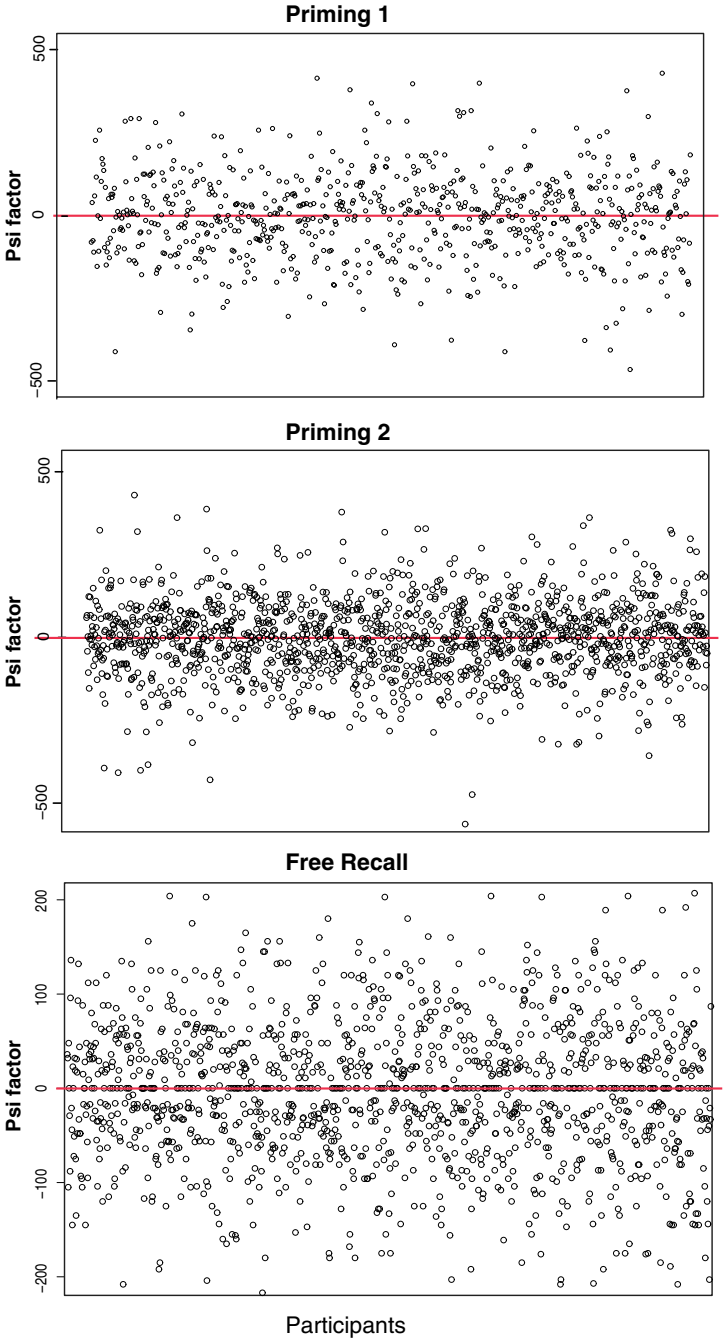
We further conducted some exploratory post hoc analyses on other variables and questionnaire items. In Experiment 1, we found a significant correlation between the item "sympathetic, warm" of the short version of the Big Five questionnaire and the *psi* factor, indicating that participants who scored high in this item showed better precognition performance. In Experiment 2, we found that participants with unusual and magical beliefs showed a significant precognition effect, a correspondence that has been already reported previously (Rattet & Bursik, 2001). In the free recall experiment, the ANOVA between substance consumption and *psi* performance was significant, indicating that regular consumers of psychoactive substances showed a precognition effect. This was a potentially interesting finding, as experiences of precognition have been reported after the use of psychoactive drugs (Luke, 2006). Moreover, two items of the short version of the Big Five questionnaire (Muck et al., 2007) were correlated with the *psi* factor: The item "anxious, easily upset" was negatively correlated and the item "calm, emotionally stable" was positively correlated with *psi* performance. The two items are strongly intercorrelated, and reflect the dimension "emotional stability" of the five-factor model, or Big Five, which is the predominant model of personality traits in psychology (Costa, 1992; Muck et al., 2007). Correlations between this dimension and *psi* performance have been reported previously (Cardeña et al., 2015).

The question thus arose whether these relationships were meaningful or not. All in all, we conducted 19 exploratory post hoc analyses, and it was therefore no surprise that some of these analyses led to significant results. When randomly selecting a number between 1 and 100 twenty times, we can expect that at least one number will be smaller than 5. The α level should therefore be corrected when several tests address a common null hypothesis, and rejection of the null hypothesis is possible when only some of the tests are found to be individually significant (Cabin & Mitchell, 2000). We therefore corrected the α level as proposed by Bonferroni ($p = 0.05/19 = .003$). After α-level correction, the only significant result was the positive correlation between supernatural beliefs and the *psi* factor in Experiment 2. This relationship was potentially interesting: Some participants could especially be sensitive for an "anomalous information transfer" underlying precognition (Bem & Honorton, 1994). This sensitivity could make them prone for *psi* experiences and thus they could have developed supernatural beliefs (Rattet & Bursik, 2001). However, extraordinary claims require extraordinary evidence and empirical data which are good enough for a priori plausible claims may not be good enough for less probable claims such as that *psi* can be scientifically proven. Even Bem stated that the requirements for evidence of precognition are "extraordinary" (Bem et al., 2015). In the present study, the correlation between supernatural beliefs and precognition did not occur systematically across experiments. Moreover statistically, precognition of those participants who scored high in supernatural belief was not significantly different of those who did not. Therefore, the finding indicating precognition must be interpreted cautiously. We propose that such potentially interesting results should be validated in further experiments using theory-based hypotheses, appropriate designs, preregistrations, and confirmatory data analyses (Goeman & Solari, 2011). For example, in order to test the possibility that supernatural beliefs are related to precognition, an experimental design could be used to compare precognition of people with high magical beliefs and people with no such beliefs directly (cf. Bem & Honorton, 1994).

Finally, we want to emphasize that the null findings for precognition in our study do not rule out the possibility that *psi* phenomena might exist in some form. The history of science has revealed many examples of phenomena that were repeatedly observed but rejected by the scientific community because they were not explainable at the time of their observation (Rabeyron, 2020). Thus, rejecting everything that is unexplainable is not justified. We should not forget that for centuries, lightning, magnetic fields, and other forms of electromagnetic radiation were unexplainable. More recently, it has been demonstrated that humans can perceive geomagnetic fields, an ability that was long disputed (Wang et al., 2019). This latter example illustrates the importance of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 4
Scatterplots Depicting the Psi Factor in All Experiments



Note. In the priming experiments, the *psi* factor depicts the difference in milliseconds between incongruent and congruent trials. In the free recall experiment, the *psi* factor depicts the number of practice words (*P*) recalled minus the number of control words (*C*) recalled multiplied by the participant's overall recall score, $(P - C) \times (P + C)$. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

a theory-driven approach in order to develop appropriate measurement methods, an issue that is largely underdeveloped in the research on *psi* phenomena.

References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Barušs, I., & Rabier, V. (2014). Failure to replicate retrocausal recall. *Psychology of Consciousness*, *1*(1), 82–91. https://doi.org/10.1037/cns0000005

Bem, D., Tressoldi, P., Rabeyron, T., & Duggan, M. (2015). Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Research*, *4*(3), Article 1188. https://doi.org/10.12688/f1000research.7177.1

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/10.1037/a0021524

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of informationtransfer. *Psychological Bulletin*, *115*(1), 4–18. https://doi.org/10.1037/0033-2909.115.1.4

Bernstein, P. (2005). Intuition: What science says (so far) about how and why intuition works. *Endophysics, Time, Quantum And The Subjective: (With CD-ROM)* (pp. 487–506). World Scientific Publishing. https://doi.org/10.1142/9789812701596_0026

Brown, J. (1968). Reciprocal facilitation and impairment of free recall. *Psychonomic Science*, *10*(2), 41–42. https://doi.org/10.3758/BF03331397

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America*, *81*(3), 246–248.

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

Cardeña, E., Palmer, J., & Marcusson-Clavertz, D. (2015). *Parapsychology: A handbook for the 21st century* (pp. 399–403). McFarland.

Cleary, A. M., & Claxton, A. B. (2018). Déjà vu: An illusion of prediction. *Psychological Science*, *29*(4), 635–644. https://doi.org/10.1177/0956797617743018

Cohn, S. A. (1999). Second sight and family history: Pedigree and segregation analyses. *Journal of Scientific Exploration*, *13*(3), 351–371.

Costa, P. (1992). *Neo PI-R professional manual lifespan and intergenerational effects of childhood malnutrition view project health correlates view project*. https://www.researchgate.net/publication/240133762

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference* (p. 44). Palgrave Macmillan.

Dienes, Z. (2021). Obtaining evidence for no effect. *Collabra. Psychology*, *7*(1), Article 2021. https://doi.org/10.1525/collabra.28202

Erickson, D. L. (2011). Intuition, telepathy, and interspecies communication: A multidisciplinary perspective. *NeuroQuantology : An Interdisciplinary Journal of Neuroscience and Quantum Physics*, *9*(1), 145–152. https://doi.org/10.14704/nq.2011.9.1.399

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Franklin, M. S., Baumgart, S. L., & Schooler, J. W. (2014). Future directions in precognition research: More research can bridge the gap between skeptics and proponents. *Frontiers in Psychology*, *5*, Article 907. https://doi.org/10.3389/fpsyg.2014.00907

Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate ψ. *Journal of Personality and Social Psychology*, *103*(6), 933–948. https://doi.org/10.1037/a0029709

Galak, J., & Nelson, N. D. (2010). *A replication of the procedures from Bem (2010, Study 8) and a failure to replicate the same results*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1699970

Gauvrit, N. (2011). Precognition or pathological science? An analysis of Daryl Bem's controversial "Feeling the Future" paper. *Skeptic*, *16*(4), 28–32.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Goeman, J. J., & Solari, A. (2011). Multiple testing for exploratory research. *Statistical Science*, *26*(4), 584–597. https://doi.org/10.1214/11-STS356

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behaviour Research and Therapy*, *34*(5–6), 489–499. https://doi.org/10.1016/0005-7967(95)00082-8

Hameroff, S. (2012). How quantum brain biology can rescue conscious free will. *Frontiers in Integrative Neuroscience*, *6*, Article 93. https://doi.org/10.3389/fnint.2012.00093

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Physics of Life Reviews, *11*(1), 39–78. https://doi.org/10.1016/j.plrev.2013.08.002

Hawking, S., & Penrose, R. (2010). *The nature of space and time*. Princeton University Press. https://doi.org/10.1515/9781400834747

Hermans, D., De Houwer, J., & Eelen, P. (2010). A time course analysis of the affective priming effect. *Cognition and Emotion*, *15*(2), 143–165. https://doi.org/10.1080/02699930125768

Honorton, C., & Ferrari, D. C. (1989). "Future telling": A meta-analyses of forced-choice precognition experiments, 1935–1987. *The Journal of Parapsychology*, *53*(4), 281–308.

Honorton, C., Ferrari, D. C., & Bem, D. J. (1998). Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology*, *62*(3), 255–276. https://www.proquest.com/docview/195033127

Houran, J., Lange, R., & Hooper, D. (2018). Cross-examining the case for precognition: Comment on Mossbridge and Radin (2018). *Psychology of Consciousness*, *5*(1), 98–109. https://doi.org/10.1037/cns0000126

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

Jolij, J., & Bierman, D. (2019). Two attempted retro-priming replications show theory-relevant anomalous connectivity. *Journal of Scientific Exploration*, *33*(1), 43–60. https://doi.org/10.31275/2019.1262

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *The Journal of Parapsychology*, *40*(1), 1–33.

Klauer, K., & Musch, J. (2003). *Affective priming: Findings and theories. The psychology of evaluation: Affective processes in cognition and emotion*. Erlbaum.

Klein, C., Andresen, B., & Jahn, T. (1997). Erfassung der schizotypen Persoenlichkeit nach DSM-III-R: Psychometrische Eigenschaften einer autorisierten deutschsprachigen Uebersetzung des ''Schizotypal Personality Questionnaire'' (SPQ) von Raine [Assessment of schizotypal personality according to DSM-III-R=Psychometric properties of an authorized German translation of ''Schizotypal Personality Questionnaire'' by Raine]. *Diagnostica*, *43*(4), 347–369.

Lang, P. J., Bradley, M. M., & Cuthbert, S. (2005). International affective picture system (IAPS) : Affective ratings of pictures and instruction manual. *Technical Report*. https://ci.nii.ac.jp/naid/20001061266

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, *15*(4), 371–379. https://doi.org/10.1037/a0025172

Luke, D. (2006). Paranormal phenomena and psychoactive drugs: Fifty years of research. *Bulletin of the Multidisciplinary Association for Psychedelic Studies*, *15*, 15–16.

Maier, M. A., Büchner, L., Kuhbandner, C., Pflitsch, M., Fernández-capo, M., & Gámiz-sanfeliu, M. (2014). Feeling the future again. *Journal of Consciousness Studies*, *21*(9–10), 121–152. https://philpapers.org/rec/MAIFTF

Marwaha, S. B., & May, E. C. (2015). Rethinking extrasensory perception: Toward a multiphasic model of precognition. *SAGE Open*, *5*(1), 1–17. https://doi.org/10.1177/2158244015576056

McMullin, E. (2008). *The virtues of a good theory*. The Routledge Companion to Philosophy of Scienceophy of Science.

Morgan, J. F. (2007). *p* Value fetishism and use of the Bonferroni adjustment. *Evidence-Based Mental Health*, *10*(2), 34–35. https://doi.org/10.1136/ebmh.10.2.34

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology*, *3*, Article 390. https://doi.org/10.3389/fpsyg.2012.00390

Mossbridge, J. A., Tressoldi, P., Utts, J., Ives, J. A., Radin, D., & Jonas, W. B. (2014). Predicting the unpredictable: Critical analysis and practical implications of predictive anticipatory activity. *Frontiers in Human Neuroscience*, *8*, Article 146. https://doi.org/10.3389/fnhum.2014.00146

Muck, P. M., Hell, B., & Gosling, S. D. (2007). Construct validation of a short five-factor model instrument: A self-peer study on the German adaptation of the Ten-Item Personality Inventory (TIPI-G). *European Journal of Psychological Assessment*, *23*(3), 166–175. https://doi.org/10.1027/1015-5759.23.3.166

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519

Rabeyron, T. (2014). Retro-priming, priming, and double testing: Psi and replication in a test–retest design. *Frontiers in Human Neuroscience*, *8*, Article 154. https://doi.org/10.3389/fnhum.2014.00154

Rabeyron, T. (2020). Why most research findings about psi are false: The replicability crisis, the psi paradox and the myth of Sisyphus. *Frontiers in Psychology*, *11*, Article 562992. https://doi.org/10.3389/fpsyg.2020.562992

Rabeyron, T., & Watt, C. (2010). Paranormal experiences, mental health and mental boundaries, and psi. *Personality and Individual Differences*, *48*(4), 487–492. https://doi.org/10.1016/j.paid.2009.11.029

Radin, D., Modestino, E. J., & Manolea, A. (2017). Electrocortical correlations between pairs of isolated people: A reanalysis. *F1000Research*, *6*, Article 676. https://doi.org/10.12688/f1000research.11537.1

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Radin, D. I. (2004). Electrodermal presentiments of future emotions. *Journal of Scientific Exploration*, *18*(2), 253–273.

Raine, A., & Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. https://doi.org/10.1093/schbul/17.4.555

Rattet, S. L., & Bursik, K. (2001). Investigating the personality correlates of paranormal belief and precognitive experience. *Personality and Individual Differences*, *31*(3), 433–444. https://doi.org/10.1016/S0191-8869(00)00148-3

Savva, L. (2014). *Is some of the evidence for ostensible precognition indicative of Darwinian adaptation to retrocausal influences? Thesis*. https://doi.org/10.25602/GOLD.00010721

Schwarzkopf, D. S. (2014). We should have seen this coming. *Frontiers in Human Neuroscience*, *8*, Article 332. https://doi.org/10.3389/FNHUM.2014.00332/TEXT

Sheehan, D. (2006). *Frontiers of tme: Retrocausation-experiment and theory*. https://philpapers.org/rec/SHEFOT-2

Voelkl, B. (2019). Multiple testing: Correcting for alpha error inflation with false discovery rate (FDR) or family-wise error rate? *Animal Behaviour*, *155*, 173–177. https://doi.org/10.1016/j.anbehav.2019.07.001

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (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. https://doi.org/10.1037/a0022790

Wang, C. X., Hilburn, I. A., Wu, D.-A., Mizuhara, Y., Cousté, C. P., Abrahams, J. N. H., Bernstein, S. E., Matani, A., Shimojo, S., & Kirschvink, J. L. (2019). Transduction of the geomagnetic field as evidenced from alpha-band activity in the human brain. *eNeuro*, *6*(2), Article ENEURO.0483-18.2019. https://doi.org/10.1523/ENEURO.0483-18.2019

Wittmann, M., Scheck, F., Feldmann, J., Glaesmann, A., Mossbridge, J., & Bem, D. (2021). The German version of a retroactive priming task shows mixed effects. *Psychology of Consciousness*. Advance online publication. https://doi.org/10.1037/cns0000296

Zingrone, N., Alvarado, C., & Dalton, K. (1999). Psi experiences and the "Big Five": Relating the NEO-PI-R to the experience claims of experimental subjects. *European Journal of Parapsychology*, *14*, 31–51.

Received February 25, 2022
Revision received August 17, 2022
Accepted August 18, 2022 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.