

# Psi Performance as a Function of Demographic and Personality Factors in Smartphone-Based Tests: Using a “SEARCH” Approach

Julia Mossbridge,<sup>1,2</sup> University of San Diego  
Dean Radin, Institute of Noetic Sciences

**Abstract:** *Objective:* We set out to gain a better understanding of human psychic or “psi” functioning by using a smartphone-based app to gather data from thousands of participants. Our expectations were that psi performance would often be revealed to be in the direction opposite to the participants’ conscious intentions (“expectation-opposing”; previously called “psi-missing”), and that gender and psi belief would be related to performance. *Method:* We created and launched three iOS-based tasks, available from 2017 to 2020, related to micro-psychokinesis (the ability to mentally influence a random number generator) and precognition (the ability to predict future randomly selected events). We statistically analyzed data from more than 2,613 unique logins and 995,995 contributed trials using null hypothesis significance testing as well as a pre-registered confirmatory analysis. *Results:* Our expectations were confirmed, and we discovered additional effects *post-hoc*. Our key findings were: 1) significant expectation-opposing effects, with a confirmatory pre-registered replication of a clear expectation-opposing effect on a micro-PK task, 2) performance correlated with psi belief on all three tasks, 3) performance on two of the three tasks related to gender, 4) men and women apparently used different strategies to perform micro-PK and precognition tasks. *Conclusions:* We describe our recommendations for future attempts to better understand performance on forced-choice psi tasks. The mnemonic for this strategy is SEARCH: Small effects, Early and exploratory, Accrue data, Recognize diversity in approach, Characterize rather than impose, and Hone in on big results.

---

1 Address correspondence to: Julia Mossbridge, Ph. D., TILT: The Institute for Love and Time, PO Box 2814 Sebastopol, CA 95472, USA, [jmossbridge@gmail.com](mailto:jmossbridge@gmail.com)

2 Dept. of Physics and Biophysics, University of San Diego; TILT: The Institute for Love and Time; Institute of Noetic Sciences



*Keywords:* precognition, psychokinesis, micro-PK, remote viewing, psychic abilities, anomalous information reception, parapsychology, anomalous cognition, psi

## Highlights

- To better understand psi, we designed a smartphone app with three “games” that test precognition (prediction of random future events) and micro-psychokinesis (micro-PK; mental influencing of random number generators).
- We performed exploratory analyses on data from more than 2,613 unique logins and 995,995 contributed trials to determine demographic and personality traits that correlate with psi performance.
- We found significant psi performance on all three games and complex relations with psi belief, gender, and personality traits.
- In a pre-registered confirmatory analysis, we confirmed the existence of a large and initially unexpected micro-PK effect.
- We discuss our approach to analyzing data from non-laboratory forced-choice psi experiments and make recommendations for future research.

From the beginning of systematic psi research, progress has been inhibited by funding issues largely resulting from an under-informed taboo and by assumptions about what effects are possible from the perspective of scientific materialism (cf. Cardeña, 2018; Franklin et al., 2018; Mossbridge & Radin, 2018a,b). In addition, three research problems have reduced the speed of discovery in the field. First, psi effects tend to be small and heterogeneous when drawing unselected test participants from the general population. This makes psi difficult to demonstrate or to replicate within typical brief laboratory experiments that rely on small numbers of participants. Second, the number of individuals required to identify participants who exhibit consistent talent on specific psi tasks is not feasible to obtain for most psi laboratories. And third, individual differences appear to affect performance on psi tasks, making it difficult to find a single “gold-standard” task that can be reliably used to probe factors that influence performance.

Fortunately, with the rise of the Internet, each of these problems can now be addressed through the use of online psi experiments (e.g., Radin, 2018). Here we report data from three smartphone tasks active from 2017 to 2020. The data were obtained from 2,613 participants (or more precisely, by that number of unique logins), who to-

gether contributed 995,995 trials. The tasks were designed to investigate micro-psychokinesis (micro-PK) and precognition. Our focus was on understanding how demographic and personality factors affected psi performance on these tasks.

To help explain our approach, consider that when the neural strategies for processing visual input were only partially understood, one key method used in visual neuroscience was to analyze correlations among perceptual skills to probe the brain's perceptual strategies (Karni & Bertini, 1997). Later, this same approach was used in auditory neuroscience (Mossbridge et al., 2006; Mossbridge et al., 2008). The idea is that if an individual were trained to improve their performance on one task, and that improvement translated into improvement on another task that was not trained, then those two tasks were likely to share a common substrate. A similar approach in psi research has been the finding that a robust meditation practice seems to have positive influences on precognition (Roney-Dougal & Solfvin, 2011; Roney-Dougal et al., 2008) and micro-PK (Braud, 1989; 2002) performance, a finding that may suggest that meditation drives changes that benefit psi. By comparison, brief alterations of consciousness using hypnotic suggestion have had no consistent effect on precognition or micro-PK tasks (Lantz, 1989; Mossbridge et al., 2021). Taken together, these results suggest that for those who are not naturally talented with psi abilities, the mechanisms responsible for improved precognition or micro-PK performance require long-term changes in neural plasticity that cannot be induced by a temporary change in one's state of consciousness.

Although it would be ideal to apply the learning-generalization approach to psi functioning, previous attempts to do so have failed due to the small and inconsistent effects obtained on the types of forced-choice psi tasks that are most easily amenable to daily training (Mossbridge et al., 2009). Thus, we took the approach of ignoring any potential learning effects and instead examined overall performance on different tasks as a function of gender, age, psi belief, confidence in psi abilities, and the Big-5 personality factors. The logic was that if any of these factors reliably related to performance on any psi task, these correlations could provide hints about mechanisms underlying task performance. Further, by examining performance across multiple types of psi tasks we might possibly gain insight into the strategies used to perform those tasks and the relations among their underlying mechanisms.

This trait-analysis approach is not new to psi and was used decades ago in an attempt to create a "signature" for psi performance (Radin, 1989). In terms of the Big-5 personality traits and their relation to psi performance, openness to experience and extraversion have been positively related to psi performance in general, especially for precognition (Hitchman et al., 2012; Honorton et al., 1998; Palmer & Carpenter,



1998; Zdrenka & Wilson, 2017). Psi belief, which is closely related to psi experience, also seems to be related to better performance on psi tasks (Braud, 2002; Lawrence, 1993; Marcusson-Clavertz & Cardeña, 2011; Palmer, 1971; Storm & Tressoldi, 2017). And when psi belief was manipulated, psi performance on a clairvoyance task was successfully manipulated as well (Walsh & Moddel, 2007). In addition, gender seems to influence both precognition and psychokinesis tasks, but in complex and inconsistent ways that do not allow firm conclusions to be drawn, except that as in many non-psi behavioral tasks men and women sometimes have different ways of responding to certain tasks and stimuli (Bierman & Scholte, 2002; Jahn et al., 2017; Lobach, 2009; Mossbridge, 2017; Mossbridge et al., 2012; Radin & Lobach, 2007; Wittmann et al., in press).

The psi performance we investigated involved an iOS-based “Psi3” smartphone app available from 2017-2020. The app presented three tasks designed to measure performance on micro-psychokinesis, conscious precognition, and unconscious precognition. Given the many factors we examined, there are many analyses that could have been performed. We took an exploratory approach on initial data, and then pre-registered confirmatory analyses on new data where we felt the exploratory effects were intriguing or robust enough to warrant it. Although a case has been made that psi functioning may be “trickster”-like (Kennedy, 2003; Maier et al., 2018; Radin, 2019), or that it cannot function consistently as a result of inherent quantum constraints on signaling (Atmanspacher & Filk, 2012), we did not adopt those assumptions. This is because of ample evidence that pre-screened and trained individuals performing free-response remote viewing experiments can access psi skills consistently at a rate above chance (May & Marwaha, 2018; Mossbridge & Radin, 2018a; Utts, 1996). So, instead of assuming that psi can never reveal its nature to us, we assumed that until recently we simply had insufficient statistical power and overly simplistic designs to demonstrate consistent psi performance on brief trials with untrained participants.

## Methods

### Experimenter Information

This was a smartphone-based study and there were no interactions between the experimenters and the participants, though it is likely that many participants knew who the experimenters were. Both experimenters had a strong belief that the data would support the existence of psi functioning.

## Data Separation by Date

We examined two batches of data both separately and together. The first batch consisted of data recorded from the launch date of the app, June 12, 2017, to midnight GMT on April 30, 2019, and the second batch consisted of data recorded from 12:01 am GMT on May 1, 2019 to midnight GMT on April 30, 2019. They were separated into two batches because: 1) preliminary data and analyses were presented at a conference in June 2019, and 2) based on that analysis we found effects for which we felt a confirmatory analysis was warranted, so the second batch allowed us to gather those data. However, the data were only analyzed separately when it was necessary to perform confirmatory analyses (see Participants).

## Random Number Generation

All games used a random number generator that drew from a truly random source. Specifically, we used the KISS07 Java algorithm XOR'd with a rapidly changing low-order output of the phone's accelerometer (second-to-fastest-changing output). Random bits generated in this way have been checked with standard randomness testing suites and deemed adequately random. The KISS07 algorithm passes the diehard test battery and has a period greater than  $10^{36}$  (<https://groups.google.com/g/comp.lang.fortran/c/5Bi8cFoYwPE/m/pSFU7NaK224J>). In any case, phone accelerometers are good sources of true randomness even in their stationary state (Voris et al., 2011), and the output of a pseudorandom process XOR'd with truly random output must necessarily be truly random.

## Procedure

Participants were required to indicate their consent using a within-app informed consent form approved by the Institutional Review Board of the Institute of Noetic Sciences (IONS\_IRB#: 01-11-17-102). After indicating consent, they were asked to complete a brief survey that asked their age, gender, belief in their psi abilities, and confidence in their psi performance. To indicate gender they moved a slider, with the default position of the slider in the middle of the range, anchored by a male symbol on the left end and a female symbol on the right end. They did not have to change the position of the slider to continue the survey. Belief in psi and confidence in psi abilities were indicated as values 1 through 5, where 5 indicated greater belief and confidence. Finally, participants were asked to take the 10-item Brief Big-5 Personality Trait Inven-

tory (Gosling et al., 2003; McCrae & John, 1992; Tupes & Christal, 1961), scored according to the method used by McCrae & John (1992). After completing these survey items, they were able to play any of the three games or none of them, at will, for as many or as few trials as they desired.

## Participants

Except where noted in the results, we included as participants all users of the Psi3 app who performed any of the three games during the periods included within the two data batches, who listed their age as between 18 and 100 years old, and whom we considered “attentive” participants. To score attentiveness, we used their responses to the Brief Big-5 Personality Trait Inventory. Participants who responded to 2 or more reverse-scored questions with the same response as they did to its opposite were not included in the analysis. For example, those who responded the same way to both “I see myself as someone who does a thorough job” and “I see myself as someone who tends to be lazy,” and also responded the same way to both “I see myself as someone who is relaxed, handles stress well” and “I see myself as someone who gets nervous easily,” were considered to not be paying attention or taking their responses seriously and were excluded from all further analyses, except those pre-registered as confirmatory.

This amounted to 2,192 unique attentive participants in the first batch ( $M_{\text{age}} = 43$ ,  $SD = 14.5$ ), and 421 participants in the second batch ( $M_{\text{age}} = 43$ ,  $SD = 12.5$ ), 131 of which were new users in the second batch. The tasks called *Heart Quest*, *Future Feelings*, and *Hidden Gurus* (described later) were played by 1,969 (first batch)/359 (second batch), 1,869/298, and 1,857/393 unique attentive participants in the first/second batches, respectively. An email address was the only requirement for registration, so it is possible that some participants used more than one email address. Further, most (290) participants who contributed data to the second batch also contributed data to the first batch. Thus, we combined all participants across the two data batches in all analyses except those used in the pre-registered confirmatory analyses.

In terms of gender demographics, we assumed that those who did move the gender slider were committed to their gender expression, but because the default was the center of the slider we cannot assume that all participants who did not move the slider were non-binary or gender non-conforming. The slider recorded values from 0 (100% male) to 1 (100% female), with few participants moving the slider all the way to the ends to indicate gender. Thus, we arbitrarily chose a cutoff of  $<0.25$  for “trending toward male gender” and  $>0.75$  for “trending toward female gender.” In the remain-

ing text, we call these participants “male” and “female.” According to these cutoffs, there were 945 and 240 (first and second batch) participants trending toward female gender and 1,003 and 40 (first and second batch) participants trending toward male gender.

## Tasks

Of the three tasks included with the smartphone app, *Heart Quest* was designed to measure micro-PK performance, *Future Feelings* to measure unconscious precognition performance, and *Hidden Gurus* to measure conscious precognition performance (Figure 1). In no case were participants required to play a full game. If they desired, participants could see their standing in a list that ranked performance in each of the three games, a method we hoped would increase motivation to perform multiple trials and create an intention to score well.

### *Micro-PK/Heart Quest*

For Heart Quest, a completed game consisted of 10 trials in which participants were instructed to try to mentally make the heart of an animated robot glow red and play a celebratory sound. On each trial, participants pressed anywhere on the screen (Figure 1, left image), whereupon one of three sound/image sets were presented: a glowing red heart and harp-like sound (+10 points), a bright red heart and a bell sound (0 points), or a dark red heart and no sound (-10 points). All images were faded in over 150 ms, held static for 1,500 ms, and then faded out over 1,500 ms for a total display duration of 2,650 ms. Sounds were concurrent with the images but sometimes ended before the display duration was complete (and we cannot assume that all participants used the app with audio turned on). The timing between trials depended on the participant’s choice as to when to press on the screen to begin the next trial.

After each screen press, two bits were gathered from the true random number generator, without respect to the location of the screen press. If the two bits matched two reference bits that were randomly selected prior to the start of the game, the trial was worth +10 points (e.g., reference bits: 01, trial bits: 01). If one bit matched and the other did not, it was worth 0 points (e.g., reference: 01, trial: 00). And if both bits differed, it was worth -10 points (e.g., reference: 01, trial: 10). Higher cumulative scores indicated more accurate matches between trials and the reference bits. Scores were reported to users after each trial, with the final score presented at the end. An explanation of the scoring procedure was available on the app.



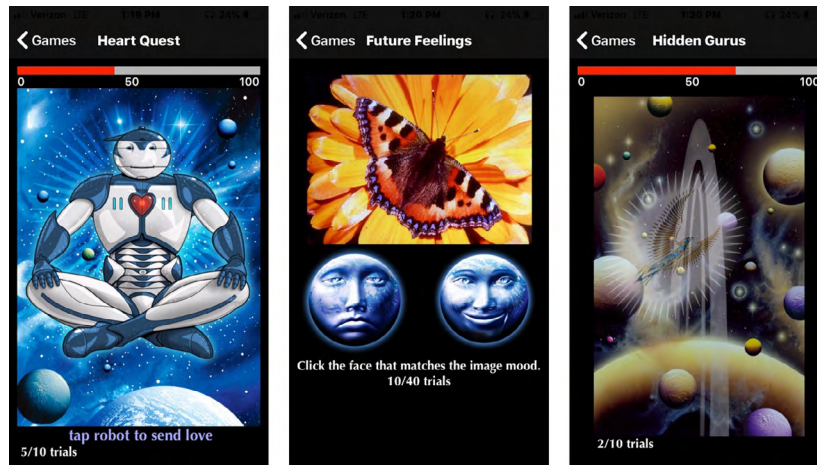
### *Unconscious Precognition/Future Feelings*

In a complete Future Feelings game, participants were instructed to respond as quickly as possible to the randomized sequential presentation of 40 target images (20 positive, 20 negative). They were instructed to press a happy face if they considered the image to be positive and a sad face if they considered the image to be negative (Figure 1, middle image). The target images were the same 40 photos as those selected from the International Affective Picture System (IAPS; Lang et al., 1997) for the laboratory version of this experiment (Bem, 2011, experiment 4). There was no time-out, that is images were displayed and stayed on screen until a response was made. However, if a response took longer than 2,500 ms, participants were shown a screen suggesting that they respond more quickly. After a response and at least a 100 ms delay, the true random number generator determined an adjective prime word they would see next on the screen; the prime was either congruent or incongruent with the valence of the image (e.g., a picture of a butterfly followed by the word “beautiful” would be a congruent pair, but a butterfly image followed by “ugly” would be incongruent). The two possible priming words associated with each target image were the same as those used by Bem (2011, experiment 4). Each priming word was presented for 1,500 ms. Then a blank screen was held for 1,500 ms prior to the presentation of the next target image. A high score indicated that correct responses were faster for congruent than incongruent pairs. Unlike in Bem’s original task, we did not play calming music or show a picture of the universe before a game was played. The user’s final score was presented at the end of a game and an explanation of the scoring was available on the app.

### *Conscious Precognition/Hidden Gurus*

In a complete Hidden Gurus game, participants tried to predict the future location of an avatar or “guru” image that would appear after the user pressed on the screen of the smartphone to make their prediction (Figure 1, right image). The location on the screen where the guru would appear on each trial was determined by the random number generator *after* the participant made their prediction. Guru images were faded on for 150 ms, then faded off over 1,500 ms. Participants determined the timing between each of 10 trials by choosing when to press on the screen to predict the next guru’s location. Scores were calculated within the app by creating an ordered list of numbers representing the distances between every pixel on the screen and the actual location of the guru, then finding the ranking on that list for the pixel representing the user’s predicted location. This method was used to create a score scaled from -10 to +10, with the final summed score provided to users at the end of the game. An explanation of the scoring was available on the app.





**Figure 1.** Screenshots of the three tasks or “games” on the Psi3 app. (Left) Heart Quest, a test of micro-psychokinesis in which participants attempted to mentally influence a random number generator to make the heart of the robot glow. (Middle) Future Feelings, a test of unconscious precognition, in which participants attempted to select a sad or happy face to reflect the valence of a target image; once selected, the image disappeared and was followed with an adjective that was congruent or incongruent with the valence of the target image just presented. (Right) Hidden Gurus, a test of conscious precognition in which participants attempted to click on the screen near where a “guru” might appear in the space-themed background.

## Analysis

### Overall

Data were analyzed in Microsoft Excel and Matlab 2018b. The threshold for statistical significance was set at  $p = 0.05$ . Two-tailed null hypothesis significance testing was used, but we report statistical measures that allow for other analyses to be performed by interested researchers. Except where noted below, or in the Results section, we considered all complete trials without regard to the number of trials required to complete each game, and we removed from analysis all incomplete trials (where a response was not made). Raw data are available upon request. All analyses were exploratory except for the pre-registered confirmatory analysis of the micro-PK effect on the micro-PK game.

### *Micro-PK/Heart Quest*

For the micro-PK task, a score for each completed trial was obtained by counting the number of reference bit pairs randomly generated at the beginning of each game

that matched with the bit pairs randomly obtained in each trial. The score was no match (0 points), partial match (1 bit matching = 1 point), or complete match (2 bits matching = 2 points). We used binomial tests to examine potential deviations from randomness in these three reference-trial bit matching levels as well as four other dependent variables: matches just to reference bit 1, reference bit 2, trial bit 1 and trial bit 2. Note that for this task we pre-registered some confirmatory analyses, and for those we did not use the “attentiveness” criterion to filter participants.

### *Unconscious Precognition/Future Feelings*

To analyze data from the unconscious precognition task, we removed from analysis all incorrect trials in which participants selected the incorrect affect for the target image, as well as correct trials with response times less than 500 ms or greater than 2,500 ms. We then categorized each trial’s response according to whether it represented a response to a positive or negative target image and whether a congruent or incongruent retro-priming word appeared after the response. This resulted in four picture-word congruence types (positive image with positive retro-prime [positive congruent]), positive image with negative retro-prime [positive incongruent], negative image with positive retro-prime [negative incongruent], and negative image with negative retro-prime [negative congruent]). This sub-categorization was necessary because there was a large bias toward responding more slowly to negative images than to positive images and we wanted to ensure that this bias did not mask any precognition effects. When we averaged response times for each participant across all trials, if there was not an averaged value for each of these four picture-word congruence types we excluded that participant’s data, as our dependent variables and their interaction required a mean value for each congruence type. These dependent variables were called  $RT_{diff_{pos}}$  (positive congruent reaction time [RT] minus positive incongruent RT) and  $RT_{diff_{neg}}$  (negative congruent RT minus negative incongruent RT).

### *Conscious Precognition/Hidden Gurus*

The dependent variable for the conscious precognition task was the same as the trial-by-trial score calculated for each participant. This score was either -10, -5, +5, or +10 (see Tasks, above), with higher scores indicating that the user chose a “guru” location closer to the future target than expected by chance.

## Results

### Micro-Pk Task/Heart Quest

#### *Trial-Level Analysis and Associated Pre-Registered Confirmatory Analyses*

Data from the Heart Quest game provided evidence for micro-PK across all eligible trials performed in both data batches. In the first batch ( $N_{\text{trials}} = 304,153$ ), there were significantly fewer trials than expected by chance that matched both reference bits, and significantly more trials than expected by chance that matched only one reference bit (first batch proportion for two matches: 0.248,  $p < 0.030$ ; proportion for one match: 0.502,  $p < 0.020$ ; proportion for no matches: 0.250,  $p > 0.600$  [binomial tests]). Thus, the first batch of data showed an expectation-opposing (i.e., psi-missing) effect, however this outcome was not replicated in the second data batch ( $N_{\text{trials}} = 70,165$ ).

We found a more interesting effect when examining the reference bits in the first batch. Although both of the bits generated for unique trials were equally likely to be 0 or 1, both of the reference bits generated once per game were significantly more likely to be a 0 rather than a 1 (Figure 2a; proportion of zeros in 1<sup>st</sup> reference bit: 0.503,  $p < 0.003$ ; 2<sup>nd</sup> reference bit: 0.506,  $p < 2 \times 10^{-10}$ ; 1<sup>st</sup> trial bit: 0.499,  $p > 0.337$ ; 2<sup>nd</sup> trial bit: 0.500,  $p > 0.955$  [binomial tests]). This was the case even though the same software function was used for generating all reference and trial bits, indicating a possible micro-PK effect (or unknown source of bias) in the randomly selected reference bits.

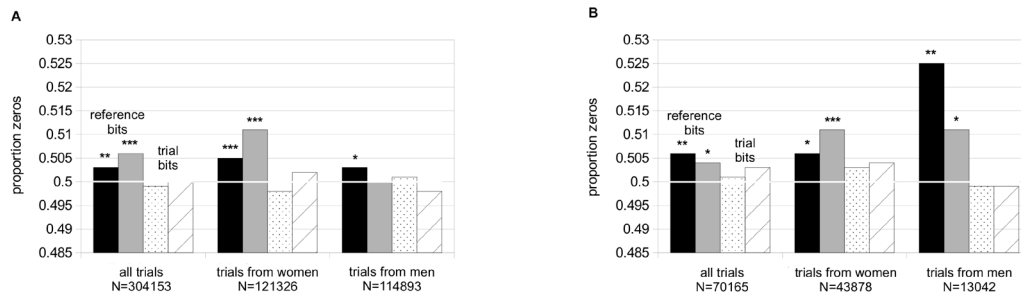
We pre-registered a confirmatory analysis of this effect with the University of Edinburgh's Koestler Unit registry prior to downloading and analyzing the data from the second batch. When we applied the same analysis to the second batch, the data revealed the same effect (Figure 2b; proportion of zeros in 1<sup>st</sup> reference bit: 0.506,  $p < 0.001$ ; 2<sup>nd</sup> reference bit: 0.504,  $p < 0.02$ ; 1<sup>st</sup> trial bit: 0.501,  $p > 0.586$ ; 2<sup>nd</sup> trial bit: 0.503,  $p > 0.147$ ). Although the power analysis specified in the pre-registration suggested we would need at least 79,000 trials to ensure an 80% chance of showing the effect and we only had 70,165 trials, the second batch revealed the same significant effects, and in the same directions, as data from the first batch, providing a clear replication.

Examining first-batch trials sorted by self-reported gender suggested that trials from individuals who reported that they were women were responsible for the apparent micro-PK effect on the reference bits. Specifically, women showed a large and significant tendency toward obtaining more zeros in both of the two reference bits (Figure 2a; first batch proportion zeros: 1<sup>st</sup> reference bit: 0.505,  $p < 0.003$ ; 2<sup>nd</sup> reference

bit: 0.511,  $p < 5 \times 10^{-15}$ ; 1st trial bit: 0.498,  $p > 0.109$ ; 2nd trial bit: 0.502,  $p > 0.240$ ). Men showed a small but significant tendency toward more zeros than ones in only the first reference bit (first batch proportion zeros: 1st reference bit: 0.503,  $p < 0.02$ ; 2nd reference bit: 0.501,  $p > 0.520$ ; 1st trial bit: 0.499,  $p > 0.727$ ; 2nd trial bit: 0.498,  $p > 0.278$ ). In data from the first batch, the proportion of zeros in the second reference bit was significantly greater for women than men (first batch  $\chi^2 = 11.55$ ;  $p < 0.0007$ ).

Given these results, we also pre-registered a confirmatory gender difference analysis to determine whether data from the second batch would replicate this effect. The confirmatory analysis did produce a significant gender difference, but in the direction opposite to that found in the first batch of data. In the second batch, reference bits from men were more likely to show the predominant-zeros effect than those from women, although women still showed the effect (Figure 2b; women, second batch proportion zeros: 1st reference bit: 0.506,  $p < 0.02$ ; 2nd reference bit: 0.511,  $p < 0.000004$ ; 1st trial bit: 0.503,  $p > 0.212$ ; 2nd trial bit: 0.505,  $p > 0.066$ ; men, second batch proportion zeros: 1st reference bit: 0.525,  $p < 6 \times 10^{-9}$ ; 2nd reference bit: 0.511,  $p < 0.02$ ; 1st trial bit: 0.499,  $p > 0.840$ ; 2nd trial bit: 0.499,  $p > 0.840$ ). This time the difference was more apparent in the first reference bit, with trials from men showing a significantly higher proportion of zeros in the first reference bit as compared to trials from women (Figure 2b; second batch  $\chi^2 = 5.16$ ;  $p < 0.03$ ).

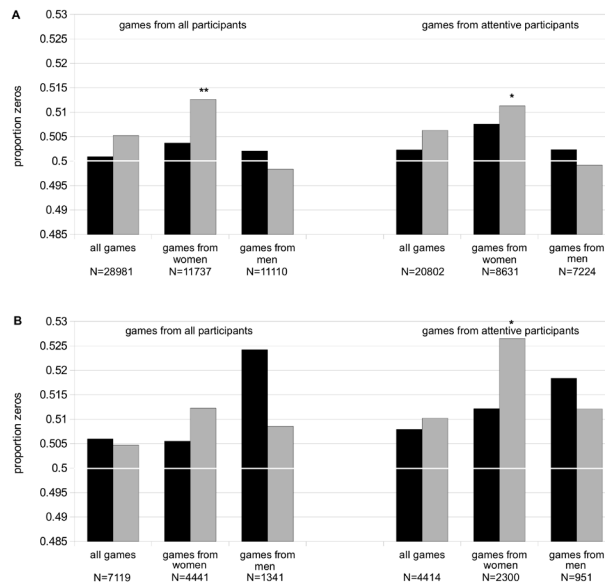
This pattern (men being responsible for the reference bit effect) was the opposite of the pattern found in the first data batch, but it is worth noting that the relation between gender and the proportion of zeros in reference bits 1 and zero was the same in both batches. That is, in both batches women had more zeros in the second than the first reference bits, while men had more zeros in the first than the second reference bits (Figure 2). Although this finding was not pre-registered, we performed chi-squared tests on the number of zeros in the first and second reference bit for men and women and found this pattern to be significant in the first batch and nearly significant in the second (first batch  $\chi^2 = 6.39$ ;  $p < 0.012$ ; second batch  $\chi^2 = 3.84$ ;  $p < 0.051$ ). Overall, some of the results observed across trials revealed significant micro-PK effects that were consistent while others differed between the two data batches; we also found a replication of an effect of gender on the relative proportion of zeros in the two reference bits.



**Figure 2.** Data from the micro-PK task for the trial-level analysis. Graphs give the proportion of zeros for both reference bits (filled bars) and both trial bits (patterned bars) for (A) trials in the first and (B) second data batches; the pre-registered analysis was performed on the second data batch. The white horizontal line across the bars at proportion = 0.5 indicates chance (zero values half the time and ones the other half). N values show the number of trials under consideration; the number of trials from men and women do not add up to the total number of trials because not all participants indicated male- or female-trending genders. Asterisks indicate significance via a binomial test vs. chance (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ ).

### Alternative Game-level Analysis

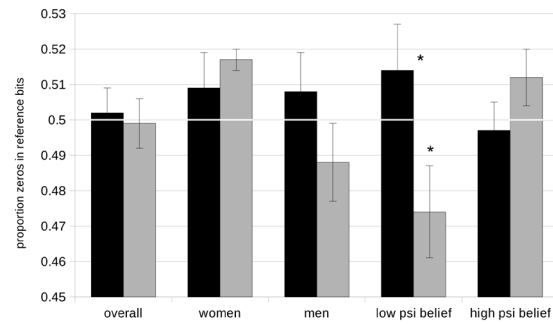
We had a concern with the pre-registered analyses in that while they allowed us to examine all trials performed by each participant and to compare non-deviations from chance among trial bits to deviations from chance among reference bits, they might provide a false impression of reference bit consistency, because reference bits are the same for every 10 trials while trial bits are not. We recognized this problem after performing the pre-registered analyses, so we could not pre-register any alternative analyses. Thus, to double-check the original results we performed an alternative analysis of reference bits at the game level (Figure 3). Within the first batch, women had significantly more zeros in the second reference bit as compared to chance, regardless of whether the games were from all women or whether we took into account only the attentive participants (as described in Methods; all women:  $p < 0.007$ ; attentive women:  $p < 0.04$  [binomial tests]). In the second batch the same pattern emerged, but it was only significant among attentive participants (proportion zeros in 2<sup>nd</sup> reference bit vs. chance: all women  $p < 0.106$ ; attentive women  $p < 0.012$  [binomial tests]). Note that the reason Figure 2 and Figure 3 do not match perfectly for reference bits is that at the game-level analysis the number of trials performed with each pair of reference bits is ignored, making reference bit pairs used in complete games (10 trials) under-represented as compared to a trial-level analysis. Importantly, the relative proportion of zeros in the two reference bits showed the same pattern across batches and matched the pattern found in the trial-level analysis (Figure 2).



**Figure 3.** Data from the micro-PK task for an alternative game-level analysis. Graphs give the proportion of zeros for both reference bits for (A) games in the first and (B) second data batches. The white horizontal line across the bars at proportion = 0.5 indicates chance. N values show the number of games under consideration. Asterisks indicate significance via a binomial test vs. chance (\* =  $p < 0.05$ , \*\* =  $p < 0.01$ ).

### Participant-level Analyses Across Both Data Batches

To examine overall and individual difference effects, we averaged data from attentive participants across all trials performed by each unique participant regardless of the data batch (first or second). Overall, there were no significant effects on the average score or on the proportion of zeros in either of the reference or trial bits (Figure 4, Table 1). Splitting the data by gender or a median split on psi belief revealed no significant effects for average score or for the average proportion of zeros in trial bits. However, with respect to the average proportion of zeros in reference bits, for individuals with low psi belief this value was both significantly lower than chance and lower than among those with high psi belief ( $t_{692} = 1.98$ ,  $p < 0.05$  vs. chance;  $t_{2185} = 2.50$ ,  $p < 0.013$  vs. high psi belief; Figure 4, Table 1). There was a tendency for women to have more zeros in the second reference bit than chance expectation and versus men, but these results were not significant ( $t_{1038} = 1.68$ ,  $p < 0.095$  vs. chance;  $t_{1924} = 1.93$ ,  $p < 0.055$  vs. men; Figure 4, Table 1). Overall, women and high psi believers showed similar patterns, while men and low psi believers showed similar patterns in terms of the relative proportion of zeros in the first and second reference bits, a pattern that matched those in Figures 2 and 3.



**Figure 4.** Results from the micro-PK task for attentive participants. Bars show mean proportion zeros for the two reference bits determined by the random function at the beginning of each game (black = 1st reference bit; gray = 2nd reference bit). The white horizontal line across the bars at proportion = 0.5 indicates chance (zero values half the time and ones the other half). Error bars give  $\pm 1$  standard error of the mean (S.E.M.). Asterisk above a bar indicates significance; asterisk between two bars indicates a significant comparison ( $* = p < 0.05$  on appropriate t-test).

**Table 1**

*Descriptive Statistics and Tests for Micro-PK Task Data from Attentive Participants*

	<b>N<sub>participants</sub></b>	<b>Score</b>	<b>1<sup>st</sup> ref. bit prop. zeros</b>	<b>2<sup>nd</sup> ref. bit prop. zeros</b>	<b>1<sup>st</sup> trial bit prop. zeros</b>	<b>2<sup>nd</sup> trial bit prop. zeros</b>
<b>Overall</b>	2187	0.997; 0.154	0.502; 0.328	0.499; 0.332	0.501; 0.102	0.501; 0.107
<b>Women</b>	1039	0.993; 0.153	0.509; 0.325	<b>0.517; 0.328; <math>p &lt; 0.095</math></b>	0.502; 0.111	0.502; 0.106
<b>Men</b>	887	1.000; 0.156	0.508; 0.331	0.488; 0.337	0.497; 0.115	0.502; 0.108
<b>Low psi belief</b>	693	1.001; 0.156	0.514; 0.346	<b>0.474; 0.348; <math>p &lt; 0.05</math></b>	0.504; 0.116	0.498; 0.113
<b>High psi belief</b>	1494	0.995; 0.153	0.497; 0.319	0.512; 0.324	0.499; 0.111	0.502; 0.103

Note: Descriptive statistics and tests against chance for the micro-PK task on averaged data from unique attentive participants. Where two values are shown, the first value is the mean, the second is the standard deviation. All comparisons versus chance giving  $p$ -values  $< 0.1$  are shown and bolded. "Score" (average proportion of 'hits') and the proportions of zeros in both trial bits are given for completeness, but they never revealed significant effects.

To further examine these patterns, we calculated an interaction term  $Ref_1 - Ref_2$  by subtracting the proportion of zeros in the second reference bit from that in the first reference bit for each participant, and then examined this term according to gender

and a median split on psi belief. There were no significant effects for gender, but for low psi believers  $Ref_1-Ref_2$  was significantly more positive than chance and also more positive than for high psi believers (low psi believers  $Ref_1-Ref_2$  vs. chance:  $t_{692}=3.81$ ,  $p<0.033$  vs. chance;  $t_{2185}=2.58$ ,  $p<0.01$  vs. high psi belief). Thus, the interaction term  $Ref_1-Ref_2$  appears to be the most informative for this task, suggesting that the psi effect was largely expressed in the differential proportion of zeros in the first and second reference bits. Further, it appears that the difference initially attributable to a gender difference might have instead reflected a difference in psi belief (also see demographic analyses, below).

We used multiple linear regression to examine potential relations among the interaction term  $Ref_1-Ref_2$  and factors of personality (Big-5) and demographics (age, gender, belief in psi, and confidence in psi abilities). Results revealed a significant overall model, and when the model was reduced to include only the most powerful predictors, they were psi belief ( $p<0.04$ ), age ( $p<0.03$ ), openness ( $p<0.07$ ), conscientiousness ( $p<0.011$ ) and extraversion ( $p<0.003$ ; Table 2), with the only negative predictor being conscientiousness. Note that gender was not included in the reduced model, suggesting that in this case psi belief was more predictive than gender, despite the gender effects described above (Figures 2 & 3). However, this difference could have been due to the participants who did not choose to report their gender (leaving the gender slider in the middle of the continuum – these participants were ignored in the group-split gender comparisons). Overall, the analysis of data from the micro-PK task provides insight into factors that may have influenced micro-PK performance in general (as discussed later), and it illustrates the complexity of expectation-opposing effects.

Table 2

*Multiple Linear Regressions on DVs for All Three Tasks with Personality and Demographic Traits as Predictors*

	DV micro-PK	DV unconscious precog	DV conscious precog
Full model	<b>adj. <math>R^2=0.009</math> <math>p&lt;0.0009</math></b>	<b>adj. <math>R^2=0.005</math> <math>p&lt;0.025</math></b>	adj. $R^2=0.007$ $p<0.128$
Openness	0.011		<b>0.095</b>
Conscientiousness	<b>-0.016</b>		
Extroversion	<b>0.016</b>	2.901	
Agreeableness			



Neuroticism		-3.58	<b>-0.082</b>
Age	<b>0.002</b>		
Gender (higher is female)		-16.45	0.204
Psi belief	<b>0.028</b>	<b>-24.6</b>	0.162
Psi confidence			-0.123
Reduced model	<b>adj. <math>R^2=0.010</math> <math>p&lt;0.00007</math></b>	<b>adj. <math>R^2=0.006</math> <math>p&lt;0.002</math></b>	<b>adj. <math>R^2=0.012</math> <math>p&lt;0.025</math></b>

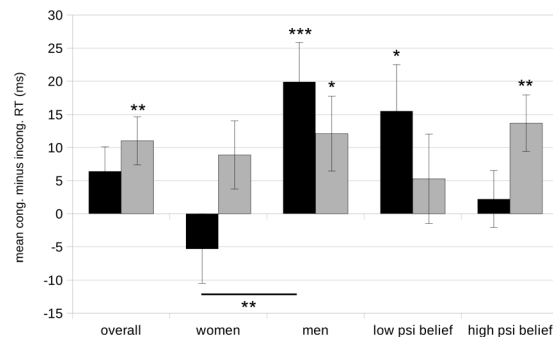
Note: Results of multiple linear regressions on averaged data from unique attentive participants who performed the micro-PK, unconscious precognition (“unconscious precog”), or the conscious precognition (“conscious precog”) task. A single multiple linear regression (full model) was performed to predict one dependent variable (DV) for each task; independent variables were personality traits and demographic traits and DVs are described in the main text (micro-PK: averaged  $Ref_1-Ref_2$ ; unconscious precognition: averaged  $RTdiff_{pos}$  minus  $RTdiff_{neg}$ ; conscious precognition: inverted averaged game score). Bold adjusted  $R^2$  values indicate significance. Estimates for each participating factor in reduced model are shown, bolded estimates indicate independent significance for that factor. Rows marked “reduced model” give results for the reduced model as a whole. Shaded cells indicate the factor was not included in the reduced model because the adjusted  $R^2$  improved when that factor was removed.

### Unconscious Precognition Task/Future Feelings

To examine the overall effects and individual differences in the *Future Feelings* task we averaged response time data for correct trials performed by attentive participants across the trials performed by each unique participant, regardless of the data batch, to obtain average response time (RT) differences  $RTdiff_{pos}$  and  $RTdiff_{neg}$ . This segregation according to target affect was critical because response times to positive images were significantly faster than those to negative images, regardless of the word primes presented after the participants’ responses (mean RT: positive target images: 1,341 ms, negative target images: 1,396 ms; positive vs. negative images  $t_{2009}=16.79$ ,  $p<1\times 10^{-58}$ ). Overall,  $RTdiff_{pos}$  and  $RTdiff_{neg}$  were both positive, indicating that for both positive and negative images, on average participants reported the target affect more quickly for targets that were incongruent with subsequent word primes as compared to congruent (Figure 5; Table 3). This was the reverse of Bem’s originally reported effect (Bem 2011, experiment 4), but the observed effect was borderline sig-

nificant for  $RTdiff_{pos}$  ( $t_{2009}=1.73$ ,  $p<0.083$ ) and significant for  $RTdiff_{neg}$  ( $RTdiff_{neg}$ :  $t_{2009}=3.04$ ,  $p<0.002$ ).

Examining the two dependent variables as a function of gender or a median split on psi belief revealed interesting effects in both cases (Figure 5, Table 3). As observed in the micro-PK task, women and men showed an inverse pattern, such that for women  $RTdiff_{pos}$  was negative while  $RTdiff_{neg}$  was positive, while for men they were both positive (interaction term  $RTdiff_{pos}$  minus  $RTdiff_{neg}$  for women versus men:  $t_{1772}=1.94$ ,  $p<0.053$ ). As would be expected from this pattern,  $RTdiff_{pos}$  was significantly higher in men as compared with women ( $t_{1772}=3.17$ ,  $p<0.0016$ ). Further, while for women only  $RTdiff_{neg}$  approached significance ( $t_{956}=1.72$ ,  $p<0.085$ ), for men both measures were significant ( $RTdiff_{pos}$ :  $t_{816}=3.34$ ,  $p<0.0009$ ;  $RTdiff_{neg}$ :  $t_{816}=2.14$ ,  $p<0.033$ ). A median split on psi belief showed that participants with high psi belief had a similar pattern to that shown by women, in that  $RTdiff_{pos}$  was more negative than  $RTdiff_{neg}$ , while an inverse pattern was apparent for participants with low psi belief (interaction term  $RTdiff_{pos}$  minus  $RTdiff_{neg}$  for low vs. high psi belief:  $t_{2008}=1.89$ ,  $p<0.059$ ). For low psi believers, only  $RTdiff_{pos}$  was significantly higher than chance, while for high psi believers, only  $RTdiff_{neg}$  was significantly higher than chance (low psi belief  $RTdiff_{pos}$ :  $t_{635}=2.23$ ,  $p<0.027$ ; high psi belief  $RTdiff_{neg}$ :  $t_{1373}=3.20$ ,  $p<0.0015$ ).



**Figure 5.** Results from the unconscious precognition task for attentive participants. Bars show the grand means for congruent minus incongruent trials for positive ( $RTdiff_{pos}$  – black bars) and negative ( $RTdiff_{neg}$  – gray bars) target images (see Methods). Chance is at zero. Error bars give  $\pm 1$  S.E.M. Asterisks above bars indicate significant versus chance; asterisks underneath line indicate significance of men vs. women comparison on  $RTdiff_{pos}$  via independent samples  $t$ -test (\* =  $p<0.05$ , \*\* =  $p<0.01$ , \*\*\* =  $p<0.001$ ).

Table 3

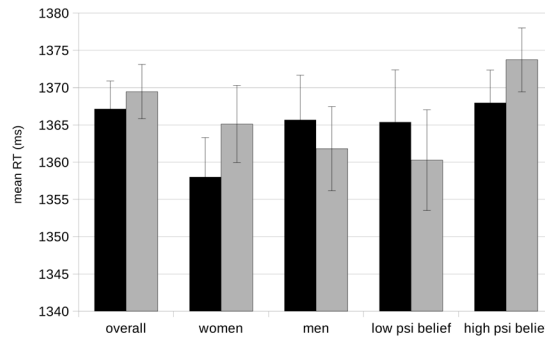
*Descriptive Statistics and Tests for Unconscious Precognition Task Data from Attentive Participants*

	N <sub>participants</sub>	RTdiff <sub>pos</sub>	RTdiff <sub>neg</sub>
Overall	2010	<b>6.425; 166.082; p&lt;0.083</b>	<b>11.034; 162.606; p&lt;0.002</b>
Women	957	-5.295; 163.794	<b>8.884; 159.424; p&lt;0.085</b>
Men	817	<b>19.894; 170.456; p&lt;0.0009</b>	<b>12.117; 161.801; p&lt;0.035</b>
Low psi belief	636	<b>15.532; 175.853; p&lt;0.027</b>	5.280; 170.783
High psi belief	1374	2.210; 161.251	<b>13.697; 158.672; p&lt;0.002</b>

Note: Descriptive statistics and tests against chance for the unconscious precognition task for averaged data from unique attentive participants (RTdiff<sub>pos</sub> = average RT for positive targets with congruent primes minus RT for positive targets with incongruent primes; RTdiff<sub>neg</sub> = same calculation for negative targets; see Methods). Where two values are shown, the first value is the mean, the second is the standard deviation. All comparisons versus chance giving *p*-values < 0.1 are shown and bolded.

The consistent trends shared by women and high psi believers on the one hand, and men and low psi believers on the other, prompted us to investigate whether these effects could have been caused by differing responses to word primes with different affects. It is common to examine the effects of congruency between a target and a prime within a priming experiment, but it is possible that for at least some participants the congruency between target and prime was not as salient as the affect of the psi cue, that is, the affect of the adjective following the response. The reason we suspected that differing responses to word primes could have been a differentiating factor is that if one type of participants (e.g., women or high psi believers) responded more swiftly to positive prime words and were less affected by congruency, then those participants would respond faster on positive-congruent trials and also faster on negative-incongruent trials (the two trial types with positive word primes), which is the pattern observed here (Figure 6). Both interactions were borderline significant versus the comparison category (for women vs. men:  $t_{1772}=1.94$ ,  $p<0.053$ ; for low vs. high psi believers:  $t_{2008}=1.90$ ,  $p<0.059$ ). Together these data suggest that the affect of the images, the congruence between images and subsequent primes, and the affect

of the primes themselves influenced performance on the unconscious precognition task. Thus, the most informative dependent variable to describe performance on this task was an interaction term ( $RT_{diff_{pos}}$  minus  $RT_{diff_{neg}}$ ).



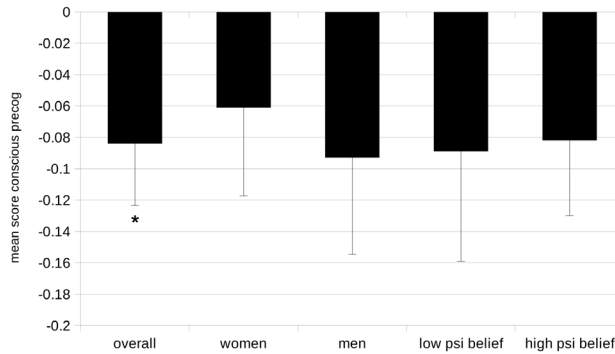
**Figure 6.** Results from the unconscious precognition task for attentive participants, differentiated by word prime affect. Bars show the grand means for response times (in ms) for positive (black bars) and negative (gray bars) word primes. Error bars give  $\pm 1$  S.E.M. Interaction comparisons between women and men and low and high psi belief showed borderline significance (see text).

Multiple linear regression and follow-up model reduction again allowed the examination of potential relations between the interaction term as the dependent variable versus all recorded demographic and personality traits as independent variables (see Methods). The overall model was significant, with the reduced model including psi belief, gender, extraversion and neuroticism (Table 2), with the only positive estimate being the one for extraversion, and with psi belief as the only independently significant predictor ( $p < 0.009$ ). Together these results support the idea that expectation-opposing effects are commonplace, and that psi belief and gender are related to performance on more than one psi task.

### Conscious Precognition Task/Hidden Gurus

The data from this conscious precognition task revealed a significant expectation-opposing effect. First, it is worth noting that for this task an alternative analysis drawing on dependent variables other than the overall score had previously been performed (Mossbridge et al., 2019). That analysis consisted of dividing the device's screen into a four-part grid and examining accuracy within each of the four quadrants. However, this had the drawback of marking as "misses" screen presses that may have been very close to the location of the future target but appearing in a neighboring quadrant. It also had the drawback of marking as "hits" screen presses within a quadrant but actually quite far away from the future target. As a result, we abandoned

that analysis and used as our dependent variable the score for each trial, averaged for all trials performed by each attentive participant across both batches. Higher scores indicated that, on average, a participant's predictions were closer to the future location of the target. Average scores were significantly lower than 0 ( $t_{2153} = -2.15, p < 0.033$ ). There were no clear additional effects when the data were separated according to gender or a median split on psi belief (Figure 7; Table 4).



**Figure 7.** Results from the conscious precognition task for attentive participants. Bars show the grand means for trial scores. Negative values indicate that participants choices were farther from the “guru” to be revealed in the future than expected by chance; chance is at zero. Error bars give  $\pm 1$  S.E.M. Asterisk below bar indicates significant versus chance ( $p < 0.05$ ).

**Table 4**

*Descriptive Statistics and Tests for Conscious Precognition Task Data from Attentive Participants*

	<b>N<sub>participants</sub></b>	<b>score</b>
Overall	2154	<b>-0.084; 1.833; <math>p &lt; 0.033</math></b>
Women	1036	-0.061; 1.814
Men	865	-0.093; 1.809
Low psi belief	690	-0.089; 1.841
High psi belief	1464	<b>-0.082; 1.830; <math>p &lt; 0.085</math></b>

Note: Descriptive statistics and tests against chance for the conscious precognition task for averaged scores from unique attentive participants. Where two values are shown, the first value is the mean, the second is the standard deviation. Any comparisons versus chance giving a  $p$ -value  $< 0.1$  is shown and bolded.



Because there was a significant expectation-opposing effect, we inverted the signs of the average scores for each participant prior to regression analyses, as the psi effect was clearly in the direction of avoiding the “hidden gurus” at a rate higher than chance. The full linear regression model on the inverted averaged scores including all recorded factors was not significant, but because the adjusted  $R^2$  value was on par with the other two tasks, we examined the reduced model. The reduced model was significant and included psi belief, gender, psi confidence, openness, and neuroticism (Table 2), with gender and psi belief as positive estimates, psi confidence and neuroticism as negative estimates, and openness and neuroticism as the only independently significant predictors ( $p < 0.03$  for both). In sum, analyses of data from the conscious precognition task underscore the complexity and task dependence of psi performance patterns, and once again support the inclusion of psi belief as a predictive factor.

### Correlations Across Performance on the Three Smartphone Tasks

Across the three tasks there were some similarities in performance patterns, especially in the conscious micro-PK task and the unconscious precognition task. However, we saw clear differences in the multiple linear regression results, indicating that the traits we examined may differently influence performance on the tasks. This suggests that performance on the tasks was governed by different mechanisms, an idea that would be supported if there were no relations between performance measures across the three tasks. To investigate this possibility, we first eliminated averaged data from attentive participants who did not contribute to the dependent variables from all three tasks. Among the 1,881 remaining participants, we found no significant correlations between these dependent variables (DV) (predicting conscious precognition DV with other two tasks as IVs:  $p > 0.237$ , adj.  $R^2 = 0.0005$ ; unconscious precognition DV with other two tasks as IVs:  $p > 0.277$ , adj.  $R^2 = 0.0003$ ; conscious micro-PK DV with other two tasks as IVs:  $p > 0.483$ , adj.  $R^2 = -0.0003$ ), apparently providing support for the idea that performance on the tasks might have been governed by at least partially independent mechanisms.

However, repeating these regression analyses on the same data but separated by gender (men alone, women alone) or by a median split on psi belief (low psi belief alone, high psi belief alone), indicated that these differentiations highlighted contrasting strategies used to perform the tasks (Table 5). For women there were no consistent or significant correlations between dependent variables from the three smartphone tasks. By contrast, men’s performance measures from the micro-PK task and the conscious precognition task were significantly correlated, suggesting that men used related strategies to perform these tasks while women did not.

For low psi believers no correlations were significant, but for high psi believers, the unconscious precognition task DV was predicted by performance on the other two tasks, and the DV from the conscious precognition task was predicted by performance on the other two tasks as well (Table 5). As might be expected based on these results, we found significant relations across task performance in men who were also high psi believers, for whom performance on the other remaining two tasks predicted performance on the micro-PK task and the conscious precognition task (Table 5). We take these results to indicate that men who were high psi believers may have used a consistent strategy to perform all three smartphone tasks, while other groups were less likely to do so.

Table 5

*Multiple Linear Regressions on DVs from All Three Tasks with DVs from Other Tasks as Predictors, Separated by Gender and Psi Belief*

	Women	Men	Low psi belief	High psi belief	Men with high psi belief
Full models					
DV micro-PK	adj. $R^2=-0.001$ $p>0.604$	<b>adj. <math>R^2=0.008</math></b> <b><math>p&lt;0.02</math></b>	adj. $R^2=-0.003$ $p>0.937$	adj. $R^2=0.0009$ $p>0.204$	<b>adj. <math>R^2=0.014</math></b> <b><math>p&lt;0.02</math></b>
DV unconscious precog	adj. $R^2=-0.0008$ $p>0.526$	adj. $R^2=-0.001$ $p>0.604$	adj. $R^2=-0.003$ $p>0.855$	<b>adj. <math>R^2=0.003</math></b> <b><math>p&lt;0.05</math></b>	adj. $R^2=0.003$ $p>0.194$
DV conscious precog	adj. $R^2=0.0003$ $p>0.317$	<b>adj. <math>R^2=0.008</math></b> <b><math>p&lt;0.03</math></b>	adj. $R^2=-0.003$ $p>0.836$	<b>adj. <math>R^2=0.004</math></b> <b><math>p&lt;0.03</math></b>	<b>adj. <math>R^2=0.014</math></b> <b><math>p&lt;0.02</math></b>
Reduced models – estimates					

IV micro-PK		<b>0.386 for DV conscious precog</b>		0.169 for DV conscious precog; 16.541 for DV unconscious precog	<b>0.488 for DV conscious precog</b>
IV unconscious precog				<b>-0.0005 for DV conscious precog</b>	-0.0005 for DV conscious precog; 0.0001 for DV micro-PK
IV conscious precog		<b>0.024 for DV micro-PK</b>		<b>-7.959 for DV unconscious precog</b>	<b>0.03 for DV micro-PK</b>

Note: Results of multiple linear regressions to examine the relations between dependent variables (DVs) from each of the three tasks. Data were averaged for each unique attentive participant who performed all three smartphone tasks in both data batches, then separated into women, men, low psi belief, and high psi belief for independent regressions. Three independent multiple linear regressions (full models) were used to predict performance on each task as the DV; independent variables (IVs) were from the remaining tasks (micro-PK: averaged Ref<sub>1</sub>-Ref<sub>2</sub>; unconscious precognition: averaged RTdiff<sub>pos</sub> minus RTdiff<sub>neg</sub>; conscious precognition: inverted averaged score). Rows marked "reduced model" give results for the reduced model where full models were significant at  $p < 0.05$ . Shaded cells mean the factor was not included in the reduced model because the adjusted  $R^2$  improved when that factor was removed. Bold indicates significance for full models and the significance of independent factors for reduced models.

**Table 6**

*Results of Multiple Linear Regression on Psi Belief with Demographic and Personality Traits as Predictors*

	DV psi belief
Full model	<b>adj. <math>R^2=0.273</math> <math>p &lt; 3 \times 10^{-16}</math></b>
Openness	-0.009
Conscientiousness	



Extroversion	
Agreeableness	<b>0.026</b>
Neuroticism	
Age	<b>0.0049</b>
Gender (higher is female)	<b>0.154</b>
Psi confidence	<b>0.368</b>
Reduced model	<b>adj. <math>R^2=0.2739</math> <math>p&lt;3\times 10^{-16}</math></b>

Note: Results of multiple linear regression examining the factors predicting psi belief among participants who performed all three tasks during the time periods of both data batches. Independent variables included all collected demographic and personality traits. Shaded cells indicate the factor was not included in the reduced model because the adjusted  $R^2$  improved when that factor was removed. Bold indicates significance for full models and the significance of independent factors for reduced models. Rows marked "reduced model" give results for the reduced model as a whole.

### Predictors of Psi Belief for Participants Performing All Three Smartphone Tasks

Psi belief was the only predictor that was consistently included in the reduced model for each smartphone task. But what is the relation between psi belief and the other traits we examined? To answer this question, we performed a full multiple linear regression on psi belief, using as predictors data from all other available demographic and personality traits. The results revealed a significant prediction of psi belief (adjusted  $R^2 = 0.274$  in the reduced model; Table 6). Psi belief was positively and independently predicted by gender ( $p < 0.000003$ ), psi confidence ( $p < 2 \times 10^{-16}$ ), age ( $p < 0.000005$ ), and agreeableness ( $p < 0.004$ ). These results indicate that these four predictors were closely associated with psi belief, so that psi belief is likely to be stronger among people who self-report that they are more female, have greater confidence in their psi abilities, are older, and are more likely to go along with what is requested of them.



## Discussion

### Expectation-Opposing Effects Are Prevalent

Across all four psi tasks, the main effects had a tendency to oppose at least the principal investigator's explicit performance expectations, sometimes significantly so. Psi performance was evident in all three tasks; the effects were just in the opposite direction from the outcomes that we believe were intended by participants, who presumably wanted to score well. This over-arching finding supports the idea that one of the reasons effect sizes for forced-choice psi tasks are generally small is that there are active psi-suppressive biases at work when conscious awareness is focused on pushing task performance in a given direction (e.g., Freedman et al., 2018; Kennedy, 2003; Rabeyron, 2020).

The results from the unconscious precognition task deserve special attention, as this was the only task in which we attempted to adapt an existing psi task protocol for online use (Bem, 2011, experiment 4). Since the original presentation of the retroactive priming effect, a meta-analysis examining 90 precognition experiments, including 15 retroactive priming experiments, revealed a significant effect in the predicted direction for the retroactive priming results, with a small effect size and relatively high heterogeneity (Bem et al., 2015). This indicates a weak effect in which responses to congruent trials were significantly *faster* than responses to incongruent trials. In the current results, responses on congruent trials were *slower* as compared to responses on incongruent trials. However, at least one laboratory examination of retroactive priming conducted following the meta-analysis described above also revealed significant effects in the reverse-congruency direction (Wittmann et al., in press).

What could cause this reversal of the canonical congruency effect? One explanation is that when participants are aware that a task is testing for an unconscious bias of some sort, and they are motivated to perform accurately, responses can reflect an unconscious compensation for the bias. This effect has been demonstrated in forward-priming experiments (Glaser & Kihlstrom, 2005; Hermans et al., 2003). However, in the latter case the authors interpreted the results to indicate that when priming stimuli are presented subliminally versus supraliminally the priming effect reverses direction (Banse, 2001). In the case of our unconscious precognition task, because they knew we were looking for an unconscious bias, it is possible that participants attempted to put themselves in a mode in which all input was regarded as sub- or trans-liminal, even though all stimuli were presented supraliminally, albeit with primes from the future. Further, in our task extreme contrasts between targets and primes were presented among

the stimuli (e.g., a picture of a toilet with excrement in it followed by the word “beautiful”). According to Glaser and Kihlstrom (2005), among highly motivated individuals performing a forward-priming task such extreme contrasts between primes and targets can produce unconscious compensation that can again produce priming effects with directions that counter expectations. Finally, we found a remarkably consistent effect in which participants responded more slowly to negative than positive target images. Using images as primes, a previous forward-priming experiment revealed that anxious individuals were significantly slower to respond to negative images (Hermans et al., 2003), suggesting that our participants may have experienced performance anxiety on this task. In any case, there were multiple elements in our unconscious precognition task that were seemingly consistent with forward-priming demonstrations of reverse-congruency effects. Future research will be necessary to determine why some retroactive priming experiments induce such expectation-opposing effects.

### **Psi Strategies Differ Across Demographic and Personality Traits**

Previous examinations of forced-choice psi task performance have provided some indications that belief in psi and personality traits such as extraversion and openness can influence accuracy, albeit in a task-specific way (Hitchman et al., 2012; Honorton et al., 1998; Marcusson-Clavertz & Cardeña, 2011; Palmer & Carpenter, 1998; Zdrenka & Wilson, 2017) and that gender or sex at birth can also have task-specific influences on psi accuracy (Bierman & Scholte, 2002; Lobach, 2009; Mossbridge, 2017; Mossbridge et al., 2012; Radin & Lobach, 2007; Wittmann et al., in press). Our exploratory conclusion after examining data from the four online forced-choice tasks described here is that the task-specificity of these factors is strongly supported.

Age influenced performance only on the micro-PK task, such that with higher age the dependent variable (the difference between the proportion of zeros in the two reference bits) increased in the same direction as it did with increases of extraversion and psi belief. Meanwhile, the most powerful predictor of performance on the unconscious precognition task was psi belief, such that with increases of psi belief the dependent variable (the difference between the congruency effects for positive minus negative targets) moved in the opposite direction as it did with increases in extraversion. Finally, performance on the conscious precognition task was most strongly predicted by openness and neuroticism, such that greater openness was related to increases and neuroticism was related to decreases in the dependent variable (inverted score). All significant regressions had very small effect sizes, suggesting that the relations between demographic and personality traits and psi task performance were relatively weak. Two of the more consistent effects were: 1) those at the bina-



ry poles of the self-identified genders/sexes showed opposing patterns in two of the three tasks (i.e., not the conscious precognition task ), and 2) psi belief was related to performance on all three tasks.

Comparisons of performance across all three tasks revealed significant correlations for men and high psi believers. Correlations between performance on different tasks can be taken as evidence that the correlated tasks draw on at least one overlapping resource or strategy (Karni & Bertini, 1997; Mossbridge et al., 2008). Given that women were more likely than men to self-report having high psi belief, this result suggested to us that both factors contributed to the strategy used for psi task performance. The most intriguing interpretation of these results is that for men the dominant strategy for conscious psi tasks might be to use micro-PK, while women and low psi believers seem to be using more diverse, task-specific approaches. The argument for this interpretation relies on our finding that for men in general, and for men with high psi belief in particular, performance on the conscious precognition task was significantly positively correlated with performance on the micro-PK task and vice versa. How might this strategy work for men? It is easy to imagine a micro-PK strategy being used to influence the location of the hidden gurus in the conscious precognition task and thereby boost users' scores. One can also imagine a precognition strategy being used to start the micro-PK task at a time when the reference bits maximized the dependent variable, but for the micro-PK task the dependent variable was not related to the score presented to the user, while for the conscious precognition task the dependent variable was exactly the score presented to the user. Thus, we think a more likely possibility is that men were using a micro-PK strategy for both the micro-PK and conscious precognition tasks.

In contrast, women showed significant psi effects on the micro-PK and unconscious precognition tasks, but their performance was not related across any of the three tasks, suggesting a more task-specific strategy. Within two tasks, women seemed to have taken different approaches from men, their results significantly contrasting with men in the micro-PK and unconscious precognition tasks. Note that for the unconscious precognition task, there is precedent for gender differences both within retroactive (Wittmann et al., in press) and forward (Gohier et al., 2011) priming experiments. Women have been shown to have greater sensitivity to negative images from the IAPS dataset, the set of images from which we drew our stimuli for the unconscious precognition task (for review see Barke et al., 2012). Supporting this idea, we saw a pattern in the results from the unconscious precognition task suggesting that women may have focused on the affective valence of the upcoming adjective prime rather than the cognitive congruency of the image and the prime. This idea is bolstered by results from Gohier and colleagues (2011) indicating that for women when word primes are used in forward-priming experiments the affective valence of the word prime drives response

time more than does congruency with the target. Thus, on the unconscious precognition task, women may have had a very different strategy than men, who showed reverse-congruency effects for both positive and negative word primes (Figures 5 and 6).

When it comes to gender or sex differences, we are inclined to agree with the conclusions stated in a study characterizing the neural correlates of creativity in a sustained attention task: in some cases, it may not be appropriate to present only average results since brain activity is so clearly differentiated between genders (Silberstein et al., 2019). There was a stark contrast between genders or sexes not only in psi task performance data on two of the three tasks (i.e., not the conscious precognition task) but also in the finding that people who rate themselves as more female are much more likely to believe in psi than those who rate themselves as more male. The relation between gender and psi belief is not a novel result (e.g., Wiseman & Watt, 2004; Wolfradt, 1997), but all of these results support the idea that the effects of gender or sex should not be ignored.

### Micro-Pk Effects May Vary Over Time

Micro-pk effects have been established to be small and especially difficult to replicate with unselected participants (Dechamps, 2019; Maier et al., 2018; Maier et al., 2020; Varvoglis & Bancel, 2015; 2016). Here, data from the micro-PK task revealed overall expectation-opposing effects on scoring as well as additional micro-PK effects that were replicated in a pre-registered confirmatory analysis. Two previous findings are especially intriguing in light of the present results. First, high trait anxiety or induced stress have both been shown to produce expectation-opposing (“psi-missing”) results on micro-PK tasks (Varvoglis & Bancel, 2015), just like we found when all trials in the first data batch were examined together. However, this effect was not pre-registered as a confirmatory analysis, and it did not replicate in the second data batch. Second, there is evidence that micro-PK as evidenced by performance on online tasks with unselected participants may follow a decline, peaking near the beginning of the task and falling to nothing by the end of task performance in a temporal pattern that is not yet well understood (Dechamps, 2019; Maier et al., 2018, 2020). The significant micro-PK effect that we found was not on the metric used to calculate participants’ scores: matches between trial and reference bits. Instead, the effect was that while the number of times “1” and “0” were produced for the two trial bits was distributed at chance levels, the two reference bits had more zeros than expected by chance. This was the case even though the same software function was used to obtain trial and reference bits. Intriguing to us is the notion that the present effect may relate to the within-experiment decline effect described for other micro-PK studies in that the reference bits were selected at the beginning of each game of 10 trials, right after the participant pressed a button to start the game.



The micro-PK effect replicated in the confirmatory analysis was on the bits chosen at the earliest possible time point in each game, perhaps suggesting some kind of temporal constraint on micro-PK. One possibility is that unselected participants who are not trained in mental focus or meditation may not be able to sustain the intention to do well throughout the course of a game. Another possibility is that micro-PK intention could build up over the course of a game and act retrocausally, creating the biggest effect for the games on which participants take the most time. It is tempting to test this hypothesis by examining the relation between game duration and scoring. This analysis reveals a massive effect that is quite impressive until one realizes that participants can choose to take as long as they like for each game. Thus, if they start out well they could be much more likely to take their time and draw out the game duration, producing a significant correlation between duration and score. Future designs might examine a potential effect of task duration and/or effort on micro-PK effects by offering a trial-pacing feature in which a trial must be performed within a short period after a signal is presented to the participant. Varying the total duration of the task across participants or across task iterations could then reveal a potential retrocausal effect.

**Confirmatory Conclusion: Micro-Pk Effects Can Be Related to Gender**

The interaction of the proportion of zeros in each of the two reference bits with gender was the most consistent effect across all three versions of the data analysis (all trials, Figure 2; all games Figure 3; all participants Figure 4). This consistency is important because one could imagine that the fact that the reference bits having significantly more zeros than expected by chance (all trials, Figure 1) could have something to do with people moving their iPhones in a different way at the beginning of a game (when the reference bits were selected) as compared to during a game. Because the random number generator was derived from an XOR-ing process between a fast-moving bit on the accelerometer and the output of a pseudorandom process, it would be difficult to explain how a consistent effect could be caused by a difference in moving the phone. In addition, the fact that the most consistent effect was due to gender further decreases the value of hypotheses relating to phone movement, because to explain these effects one would have to hypothesize that at the beginning of a game self-identified women moved their phones differently than self-identified men did, and this movement difference consistently resulted in output that survived XOR-ing with a completely independent pseudorandom output.

Finally, it is worth briefly pointing out that at the game-level analysis women showed the only significant effect (Figure 3), while at the trial-level analysis overall data and data from men showed a significant effect as well (Figure 2). This is likely

due to a reduction in degrees of freedom (or statistical power) between the trial- and game-level analysis. Men performed many fewer trials and games than women on this task. Future examinations of gender effects on online micro-PK tasks might do well to include pre-determined numbers of participants for each gender.

### Recommendations for Future Research

Our approach to analyzing these data was based on decades of psi research performed by psychologists and physiologists and it bore fruit here in terms of improving our understanding of forced-choice psi performance. We have created an acronym, "SEARCH," to help other experimenters remember the key points of this approach.

**Small effects:** We know that performance on forced-choice psi tasks produces small effects, so let us not expect big ones.

**Early and exploratory:** We are at the early stages of understanding what influences psi performance, so we need to do a lot of exploratory work before deciding that psi is trying to "trick" us.

**Accrue data:** Large numbers of participants give us the statistical power to observe influences on psi performance and within the last two decades it has become relatively easy to gather forced-choice data from a large number of participants.

**Recognize diversity in approach:** Multiple strategies as well as conscious and unconscious biases influence psi performance and they operate in distinct ways on different tasks for different participants.

**Characterize, do not impose:** Understanding the strategies used for each task requires determining the most psi-informative measure for that task, so imposing ideas about expected performance and its directionality is not productive. Two-tailed tests are therefore always in order for null hypothesis significance testing, at least in exploratory work.

**Hone in on big results:** Conduct pre-registered confirmatory analyses to determine if larger psi effects found in exploratory analyses are replicable.

It is particularly important to consider that these SEARCH recommendations are most useful when used in combination. For instance, the "A" in SEARCH ("Accrue data") can result in very large datasets such as ours, increasing the possibility of spurious cor-



relations causing Type 1 interpretive errors (Calude & Longo, 2017; Gandomi & Haider, 2015). However, when accruing large datasets is followed by characterizing and honing in on the found effects, meaningful anomalous effects can be confirmed. In the present case, we accrued almost 1 million trials and therefore reasonably expected some significant effects, but the most convincing results were the micro-PK effects confirmed in a pre-registered analysis and the significant relations between psi performance and belief, gender, and extraversion, which matched the results of previous studies.

Future efforts in this field are likely to employ more sophisticated and engaging tasks than used here. However, using the SEARCH analysis approach will go a long way towards determining the many factors that influence psi performance, as well as the strategies and mechanisms that correspond to these factors. Over time, these observations will allow the scientific community to further understand the conscious and unconscious mechanisms underlying performance on psi tasks.

### Author Contributions

Author 1 co-designed the mobile app, organized the data collection, carried out the statistical analyses, drafted the first version of the manuscript, and read and approved the final manuscript. Author 2 helped secure the initial funding for this project. co-designed the mobile app, was involved in revising the manuscript, and read and approved the final manuscript.

### Acknowledgments

The authors would like to thank the Bial Foundation for generously funding the authors during the course of this study (Bial grants 2014\_260, 97\_16). It is an honor to thank the Hummingbird Fund, Maureen Pelton and Charlie Hartwell, and Geoff and Leslie Oelsner, whose donations supported the further development of the Psi<sup>3</sup> app. We would also like to thank Arnaud Delorme, Greg Travis, David Micksch, Cameron Boehmer, and Mikey Siegel who helped design and code the Psi<sup>3</sup> iPhone app. We are grateful to all the volunteers and donors who provided support, as well as the attendees of the Consciousness Hacking-IONS “hackations” who helped design and code portions of the Psi<sup>3</sup> app.

### References

Atmanspacher, H., & Filk, T. (2012). Contra classical causality violating temporal bell inequalities in mental systems. *Journal of Consciousness Studies*, 19, 95–116.



- Banase, R. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition & Emotion, 15*(4), 501-520. Doi: 10.1080/02699930126251
- Barke, A., Stahl, J., & Kröner-Herwig, B. (2012). Identifying a subset of fear-evoking pictures from the IAPS on the basis of dimensional and categorical ratings for a German sample. *Journal of Behavior Therapy and Experimental Psychiatry, 43*(1), 565-572. Doi: 10.1016/j.jbtep.2011.07.006
- Barušs, I. & Mossbridge, J. A. (2016). *Transcendent mind: Rethinking the science of consciousness*. American Psychological Association. Doi: 10.1037/15957-009
- 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. *F1000Res, 4*:1188. Doi: 10.12688/f1000research.7177.2
- 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. Doi: 10.1037/a0021524
- Bierman, D., & Scholte, H. (2002). A fMRI brain imaging study of presentiment. *Journal of the International Society of Life Information Sciences, 20*(2), 380-388.
- Braud, W. G. (1989). Meditation and psychokinesis. *Parapsychology Review, 20*, 9-11.
- Braud, W. G. (2002). Psi-favorable conditions. In V. G. Rammohan (Ed.), *New frontiers of human science: A festschrift for K. Ramakrishna* (pp. 95-118). McFarland.
- Calude, C. S., & Longo, G. (2017). The deluge of spurious correlations in big data. *Foundations of Science, 22*(3), 595-612. Doi: 10.1007/s10699-016-9489-4
- Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: A review. *American Psychologist, 73*(5), 663-677. Doi: 10.1037/amp0000236
- Cieslarová, R. (2018). *Analysis of the Linux random number generator in virtualized environment*. Masaryk University Faculty of Informatics.
- Dechamps, M. (2019). *Mind-matter interactions and their reproducibility* (Doctoral dissertation, Ludwig Maximilians University, Munich).
- 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*, 907. Doi: 10.3389/fpsyg.2014.00907
- Freedman, M., Binns, M., Gao, F., Holmes, M., Roseborough, A., Strother, S., ... & Black, S. E. (2018). Mind-matter interactions and the frontal lobes of the brain: A novel neurobiological model of psi inhibition. *Explore, 14*(1), 76-85. Doi: 10.1016/j.explore.2017.08.003
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144. Doi: 10.1016/j.ijinfomgt.2014.10.007
- Glaser, J., & Kihlstrom, J. F. (2005). Compensatory automaticity: Unconscious volition is not an oxymoron. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconsciousness*. Oxford University Press. Doi: 10.1093/acprof:oso/9780195307696.003.0008

- Gohier, B., Senior, C., Brittain, P. J., Lounes, N., El-Hage, W., Law, V., ... & Surguladze, S. A. (2013). Gender differences in the sensitivity to negative stimuli: Cross-modal affective priming study. *European Psychiatry, 28*(2), 74–80. Doi: 10.1016/j.eurpsy.2011.06.007
- Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. Doi: 10.1016/s0092-6566(03)00046-1
- Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology, 57*, 97–114. Doi: 10.1037/h0087416
- Hitchman, G. A., Roe, C. A., & Sherwood, S. J. (2012). A reexamination of nonintentional precognition with openness to experience, creativity, psi beliefs, and luck beliefs as predictors of success. *Journal of Parapsychology, 76*(1), 109–145.
- Honorton, C., Ferrari, D. G. & Bem, D. J. (1998). Extraversion and ESP performance: A meta-analysis and a new confirmation. *Journal of Parapsychology, 62*(3), 255–276.
- Jahn, R. G., Dunne, B. J., Nelson, R. D., Dobyys, Y. H., & Bradish, G. J. (1997). Correlations of random binary sequences with pre-stated operator intention: A review of a 12-year program. *Journal of Scientific Exploration, 11*(3), 345–367. Doi: 10.1016/j.explore.2007.03.009
- Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology, 7*(4), 530–535. Doi: 10.1016/s0959-4388(97)80033-5
- Kennedy, J. E. (2003). The capricious, actively evasive, unsustainable nature of psi: A summary and hypotheses. *Journal of Parapsychology, 67*(1), 53–74.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International affective picture system (IAPS): Technical manual and affective ratings*. NIMH Center for the Study of Emotion Doi: 10.1037/t66667-000
- Lawrence, T. R. (1993). Gathering in the sheep and goats: A meta-analysis of forced-choice sheep-goat ESP studies, 1947–1993. *Proceedings of Presented Papers: The Parapsychological Association 36th Annual Convention, 75–86*.
- Lobach, E. (2009). Presentiment research: Past, present and future. In C. A. Roe, L. Coly, & W. Kramer (Eds) *Utrecht II: Charting the future of parapsychology* (pp. 22–45). Parapsychology Foundation.
- Maier, M. A., Buechner, V. L., Dechamps, M. C., Pflitsch, M., Kurzrock, W., Tressoldi, P., ... & Martsinkovskaja, T. (2020). A preregistered multi-lab replication of Maier et al. (2014, Exp. 4) testing retroactive avoidance. *PLoS One, 15*(8), e0238373. Doi: 10.1371/journal.pone.0238373
- Maier, M. A., Dechamps, M. C., & Pflitsch, M. (2018). Intentional observer effects on quantum randomness: A Bayesian analysis reveals evidence against micro-psychokinesis. *Frontiers in Psychology, 9*, 379. Doi: 10.3389/fpsyg.2018.00379
- Marcusson-Clavertz, D. & Cardeña, E., (2011). Hypnotizability, alterations in consciousness, and other variables as predictors of performance in a ganzfeld psi task. *Journal of Parapsychology, 75*(2), 235–259.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

- Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. *Journal of Neuroscience*, 26(49), 12708-12716. Doi: 10.1523/jneurosci.2254-06.2006
- Mossbridge, J., Graboweky M.F., & Suzuki S. (2009). Evidence for subconscious but not conscious psi in remote stare detection and precognition tasks. *Proceedings of the Parapsychological Association, 52nd Meeting*, 13-14.
- Mossbridge, J., & Radin, D. (2020). Three confirmatory analyses of precognition and micro-pk data gathered using online methods. [http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU\\_Registry\\_1056.pdf](http://www.koestler-parapsychology.psy.ed.ac.uk/Documents/KPU_Registry_1056.pdf) (retrieved 3/8/21).
- Mossbridge, J. (2017). Examining the nature of retrocausal effects in biology and psychology. *AIP Conference Proceedings 1841*, 030004. Doi: 10.1063/1.4982775
- Mossbridge, J. A., & Radin, D. (2018a). Precognition as a form of prospection: A review of the evidence. *Psychology of Consciousness: Theory, Research, and Practice*, 5(1), 78. Doi: 10.1037/cns0000121
- Mossbridge, J. A., & Radin, D. (2018b). Plausibility, statistical interpretations, physical mechanisms and a new outlook: Response to commentaries on a precognition review. *Psychology of Consciousness: Theory, Research, and Practice*, 5(1), 110-116. Doi: 10.1037/cns0000152
- Mossbridge, J., Tressoldi, P. E., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: A meta-analysis. *Frontiers in Psychology*, 3, 390. Doi: 10.3389/fpsyg.2012.00390
- Mossbridge, J. A., Nisam, M., & Crabtree, A. (2021). Can hypnotic suggestion induce feelings of unconditional love and supernormal performance?. *Spirituality in Clinical Practice*, 81(1), 30-50. Doi: 10.1037/scp0000239
- Mossbridge, J.A., Boccuzzi, M., Nisam, M., & Crabtree, A. (2019). Screening for precognition, remote viewing and micro-psychokinetic skills: Results and lessons from online psi-testing methods. *Presented at the 38th Annual Meeting of the Society for Scientific Exploration*. <https://www.youtube.com/watch?v=9imsfGdEnWo> [Retrieved 8/2/21].
- Palmer, J. (1971). Scoring in ESP tests as a function of ESP. Part 1: The sheep-goat effect. *Journal of the American Society for Psychical Research*, 65, 373-408.
- Palmer, J., & Carpenter, J.C. (1998). Comments on the extraversion-ESP meta-analysis by Honorton, Ferrari and Bem. *Journal of Parapsychology*, 62(3), 277-282.
- 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, 2468. Doi: 10.3389/fpsyg.2020.562992
- Radin, D. I. (1989). Searching for "signatures" in anomalous human-machine interaction data: A neural network approach. *Journal of Scientific Exploration*, 3(2), 185-200.
- Radin, D. (2019). Tricking the trickster: Evidence for predicted sequential structure in a 19-year on-line psi experiment. *Journal of Scientific Exploration*, 33(4), 549-568. Doi: 10.31275/2019/1429

- Radin, D., and Lobach, E. (2007). Toward understanding the placebo effect: Investigating a possible retrocausal factor. *Journal of Alternative and Complementary Medicine*, 13, 733–739. Doi: 10.1089/acm.2006.6243
- Roney-Dougal, S. M., & Solfvn, J. (2011). Exploring the relationship between Tibetan meditation attainment and precognition. *Journal of Scientific Exploration*, 25(1), 29–46.
- Roney-Dougal, S. M., Solfvn, J., & Fox, J. (2008). An exploration of degree of meditation attainment in relation to psychic awareness with Tibetan Buddhists. *Journal of Scientific Exploration*, 22(2), 161–78.
- Silberstein, R., Camfield, D. A., Nield, G., & Stough, C. (2019). Gender differences in parieto-frontal brain functional connectivity correlates of creativity. *Brain and behavior*, 9(2), e01196. Doi: 10.1002/brb3.1196
- Storm, L., & Tressoldi, P. E. (2017). Gathering in more sheep and goats: A meta-analysis of forced-choice sheep-goat ESP studies, 1994–2015. *Journal of the Society for Psychical Research*, 81(2), 79–107.
- Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings (No. ASD-TR-61-97)*. Personnel Research Lab. Doi: 10.21236/ad0267778
- Utts, J. (1995). An assessment of the evidence for psychic functioning. *Journal of Parapsychology*, 59, 289–320.
- Varvoglis, M., & Bancel, P. A. (2015). Micro-psychokinesis. In E. Cardeña, J. Palmer & D. Marcusen-Clavertz, (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 266–281). McFarland.
- Varvoglis, M., & Bancel, P. A. (2016). Micro-psychokinesis: Exceptional or universal? *Journal of Parapsychology*, 80(1), 37–44.
- Voris, J., Saxena, N., & Halevi, T. (2011). Accelerometers and randomness: Perfect together. *Proceedings of the fourth ACM conference on Wireless network security*, 115–126. Doi: 10.1145/1998412.1998433
- Walsh, K., & Moddel, G. (2007). Effect of belief on psi performance on a card guessing task. *Journal of Scientific Exploration*, 21(3), 501–510.
- Wiseman, R., & Watt, C. (2004). Measuring superstitious belief: Why lucky charms matter. *Personality and Individual Differences*, 37(8), 1533–1541. Doi: 10.1016/j.paid.2004.02.009
- Wittmann M, Scheck F, Feldmann J, Glaesmann A, Mossbridge J, & Bem D (in press). The German version of a retroactive priming task shows mixed effects. *Psychology of Consciousness: Theory, Research, and Practice*. Doi:10.1037/cns0000296
- Wolfradt, U. (1997). Dissociative experiences, trait anxiety and paranormal beliefs. *Personality and Individual Differences*, 23(1), 15–19. Doi: 10.1016/s0191-8869(97)00043-3
- Zdrenka, M., & Wilson, M. S. (2017). Individual difference correlates of psi performance in forced-choice precognition experiments: A meta-analysis (1945–2016). *Journal of Parapsychology*, 81(1), 9–32.