Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

© 2011 American Psychological Association

0022-3514/11/$12.00 DOI: 10.1037/a0021524

# Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size ($d$) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

*Keywords:* psi, parapsychology, ESP, precognition, retrocausation

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. The term is purely descriptive; it neither implies that such phenomena are paranormal nor connotes anything about their underlying mechanisms. Alleged psi phenomena include *telepathy*, the apparent transfer of information from one person to another without the mediation of any known channel of sensory communication; *clairvoyance* (sometimes called *remote viewing*), the apparent perception of objects or events that do not provide a stimulus to the known senses; *psychokinesis*, the apparent influence of thoughts or intentions on physical or biological processes; and *precognition* (conscious cognitive awareness) or *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process.

Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports nine experiments designed to test for such retroactive influence by "time-reversing" several well-established psychological effects, so that the individual's responses are obtained before the putatively causal stimulus events occur.

Psi is a controversial subject, and most academic psychologists do not believe that psi phenomena are likely to exist. A survey of 1,100 college professors in the United States found that psychologists were much more skeptical about the existence of psi than were their colleagues in the natural sciences, the other social sciences, or the humanities (Wagner & Monnet, 1979). In fact, 34% of the psychologists in the sample declared psi to be impossible, a view expressed by only 2% of all other respondents. Although our colleagues in other disciplines would probably agree with the oft-quoted dictum that "extraordinary claims require extraordinary evidence," we psychologists are more likely to be familiar with the methodological and statistical requirements for sustaining such claims and aware of previous claims that failed either to meet those requirements or to survive the test of successful replication. Several other reasons for our greater skepticism are discussed by Bem and Honorton (1994, pp. 4–5).

There are two major challenges for psi researchers, one empirical and one theoretical. The major empirical challenge, of course, is to provide well-controlled demonstrations of psi that can be replicated by independent investigators. That is the major goal in the research program reported in this article. Accordingly, the

This article was published Online First January 31, 2011.

I am grateful to the students who served as head research assistants and laboratory coordinators for their enthusiasm and dedication to this controversial enterprise: Ben Edelman, Rebecca Epstein, Dan Fishman, Jamison Hahn, Eric Hoffman, Kelly Lin, Brianne Mintern, Brittany Terner, and Jade Wu. I am also indebted to the 30 other students who served as friendly and reliable experimenters over the course of this research program. Dean Radin, senior scientist at the Institute of Noetic Sciences (IONS), and David Sherman, professor of psychology at the University of California, Santa Barbara, provided valuable guidance in the preparation of this article.

Correspondence concerning this article should be addressed to Daryl J. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853. E-mail: d.bem@cornell.edu

experiments have been designed to be as simple and transparent as possible, drawing participants from the general population, requiring no instrumentation beyond a desktop computer, taking less than thirty minutes per session, and requiring statistical analyses no more complex than a *t* test across sessions or participants.

The major theoretical challenge for psi researchers is to provide an explanatory theory for the alleged phenomena that is compatible with physical and biological principles. Although the current absence of an explanatory theory for psi is a legitimate rationale for imposing the "extraordinary" requirement on the evidence, it is not, I would argue, sufficient reason for rejecting all proffered evidence a priori. Historically, the discovery and scientific exploration of most phenomena have preceded explanatory theories, often by decades or even centuries. The major focus of this article is empirical, but I return to a brief discussion of theory at the end.

As noted above, the experiments in this article are concerned with apparent retroactive influence, a generalized form of precognition. Experimental tests of precognition have been reported for more than half a century. Most of the early experiments used forced-choice designs in which participants were explicitly challenged to guess which one of several potential targets would be randomly selected at a later time. Typical targets have been ESP card symbols, an array of colored lightbulbs, the faces of a die, or visual elements in a computer display. When a participant correctly selects the actual target-to-be, it is designated as a hit, and psi performance is typically expressed as the hit rate, the percentage of hits over trials.

A meta-analysis of all forced-choice precognition experiments appearing in English-language journals between 1935 and 1977 was published by Honorton and Ferrari (1989). Their analysis included 309 experiments conducted by 62 different investigators and involving more than 50,000 participants. Honorton and Ferrari reported a small but consistent and highly significant hit rate (mean $z = 0.69$, combined $z = 12.14$, $p = 6 \times 10^{-27}$). They also concluded that this overall result was unlikely to be significantly inflated by the selective reporting of positive results (the so-called file-drawer effect): There would have to be 46 unreported studies averaging null results for every reported study in the meta-analysis to reduce the overall significance of the database to nonsignificance.

Just as research in cognitive social psychology has increasingly pursued the study of cognitive and affective processes that are not accessible to conscious awareness and control (Bargh & Ferguson, 2000), research in psi has followed the same path, moving from explicit forced-choice guessing tasks to experiments using subliminal stimuli and implicit, indirect, or physiological responses. The trend is exemplified by several recent "presentiment" experiments, pioneered by Radin (1997), in which physiological indices of participants' emotional arousal were monitored as participants viewed a series of pictures on a computer screen. Most of the pictures were emotionally neutral, but a highly arousing negative or erotic image was displayed on randomly selected trials. As expected, strong emotional arousal occurred when these images appeared on the screen, but the remarkable finding is that the increased arousal was observed to occur a few seconds before the picture appeared, before the computer had even selected the picture to be displayed. The presentiment effect has also been demonstrated in an fMRI experiment that monitored brain activity (Bierman & Scholte, 2002) and in experiments using bursts of noise rather than visual images as the arousing stimuli (Spottiswoode & May, 2003). A review of presentiment experiments prior to 2006 can be found in Radin (2006, pp. 161–180). Although there has not yet been a formal meta-analysis of presentiment studies, there have been 24 studies with human participants through 2009, of which 19 were in the predicted direction and about half were statistically significant. Two studies with animals were both positive, one marginally and the other substantially so (D. I. Radin, personal communication, December 20, 2009).

Most of the experiments reported in this article are also part of this trend toward using subliminal stimulus presentations and indirect or implicit response measures. Each of them modified a well-established psychological effect by reversing the usual sequence of events, so that the individual's responses were obtained before rather than after the stimulus events occurred. Table 1 provides an overview of the effects and their corresponding time-reversed experiments.

## Precognitive Approach and Avoidance

The presentiment studies provide evidence that our physiology can anticipate unpredictable erotic or negative stimuli before they occur. Such anticipation would be evolutionarily advantageous for reproduction and survival if the organism could act instrumentally to approach erotic stimuli and avoid negative stimuli. The two experiments in this section were designed to test whether individuals can do so.

## Experiment 1: Precognitive Detection of Erotic Stimuli

As noted above, most of the earlier experiments in precognition explicitly challenged participants to guess which one of several stimuli would be randomly selected after they recorded their guess. In most of these experiments, participants were also given explicit trial-by-trial feedback on their performance. This first experiment adopts this traditional protocol, using erotic pictures as explicit reinforcement for correct "precognitive" guesses.

### Method

One hundred Cornell undergraduates, 50 women and 50 men, were recruited for this experiment through the Psychology Depart-

Table 1
*Overview of Psychological Effects and Their Corresponding Time-Reversed Experiments*

| Standard psychological effect | Experiments |
|---|---|
| Approach/avoidance | 1. Precognitive Detection of Erotic Stimuli |
| | 2. Precognitive Avoidance of Negative Stimuli |
| Affective priming | 3. Retroactive Priming I |
| | 4. Retroactive Priming II |
| Habituation | 5. Retroactive Habituation I |
| | 6. Retroactive Habituation II |
| | 7. Retroactive Induction of Boredom |
| Facilitation of recall | 8. Retroactive Facilitation of Recall I |
| | 9. Retroactive Facilitation of Recall II |

ment's automated online sign-up system.[1] They either received one point of experimental credit in a psychology course offering that option or were paid $5 for their participation. Both the recruiting announcement and the introductory explanation given to participants upon entering the laboratory informed them that

> this is an experiment that tests for ESP. It takes about 20 minutes and is run completely by computer. First you will answer a couple of brief questions. Then, on each trial of the experiment, pictures of two curtains will appear on the screen side by side. One of them has a picture behind it; the other has a blank wall behind it. Your task is to click on the curtain that you feel has the picture behind it. The curtain will then open, permitting you to see if you selected the correct curtain. There will be 36 trials in all.
>
> Several of the pictures contain explicit erotic images (e.g., couples engaged in nonviolent but explicit consensual sexual acts). If you object to seeing such images, you should not participate in this experiment.

The participant then signed a consent form and was seated in front of the computer. After responding to two individual-difference items (discussed below), the participant had a 3-min relaxation period during which the screen displayed a slowly moving Hubble photograph of the starry sky while peaceful new-age music played through stereo speakers. The 36 trials began immediately after the relaxation period.

**Stimuli.** Most of the pictures used in this experiment were selected from the International Affective Picture System (IAPS; Lang & Greenwald, 1993), a set of 820 digitized photographs that have been rated on 9-point scales for valence and arousal by both male and female raters. This is the same source of pictures used in most presentiment studies. Each session of the experiment included both erotic and nonerotic pictures randomly intermixed, and the main psi hypothesis was that participants would be able to identify the position of the hidden erotic picture significantly more often than chance (50%).

The hit rate on erotic trials can also be compared with the hit rates on the nonerotic trials to test whether there is something unique about erotic content in addition to its positive valence and high arousal value. For this purpose, 40 of the sessions comprised 12 trials using erotic pictures, 12 trials using negative pictures, and 12 trials using neutral pictures. The sequencing of the pictures and their left/right positions were randomly determined by the programming language's internal random function. The remaining 60 sessions comprised 18 trials using erotic pictures and 18 trials using nonerotic positive pictures with both high and low arousal ratings. These included eight pictures featuring couples in romantic but nonerotic situations (e.g., a romantic kiss, a bride and groom at their wedding). The sequencing of the pictures on these trials was randomly determined by a randomizing algorithm devised by Marsaglia (1997), and their left/right target positions were determined by an Araneus Alea I hardware-based random number generator. (The rationale for using different randomizing procedures is discussed in detail below.)

Although it is always desirable to have as many trials as possible in an experiment, there are practical constraints limiting the number of critical trials that can be included in this and several others experiments reported in this article. In particular, on all the experiments using highly arousing erotic or negative stimuli, a relatively large number of nonarousing trials must be included to permit the participant's arousal level to "settle down" between critical trials. This requires including many trials that do not contribute directly to the effect being tested.

In our first retroactive experiment (Experiment 5, described below), women showed psi effects to highly arousing stimuli but men did not. Because this appeared to have arisen from men's lower arousal to such stimuli, we introduced different erotic and negative pictures for men and women in subsequent studies, including this one, using stronger and more explicit images from Internet sites for the men. We also provided two additional sets of erotic pictures so that men could choose the option of seeing male–male erotic images and women could choose the option of seeing female–female erotic images.[2]

From the participants' point of view, this procedure appears to test for clairvoyance. That is, participants were told that a picture was hidden behind one of the curtains, and their challenge was to guess correctly which curtain concealed the picture. In fact, however, neither the picture itself nor its left/right position was determined until after the participant recorded his or her guess, making the procedure a test of detecting a future event (i.e., a test of precognition).

## Results and Discussion

Across all 100 sessions, participants correctly identified the future position of the erotic pictures significantly more frequently than the 50% hit rate expected by chance: 53.1%, $t(99) = 2.51$, $p = .01$, $d = 0.25$.[3] In contrast, their hit rate on the nonerotic pictures did not differ significantly from chance: 49.8%, $t(99) = -0.15$, $p = .56$. This was true across all types of nonerotic pictures: neutral pictures, 49.6%; negative pictures, 51.3%; positive pictures, 49.4%; and romantic but nonerotic pictures, 50.2%. (All $t$ values $< 1$.) The difference between erotic and nonerotic trials was itself significant, $t_{\text{diff}}(99) = 1.85$, $p = .031$, $d = 0.19$. Because erotic and nonerotic trials were randomly interspersed in the trial sequence, this significant difference also serves to rule out the possibility that the significant hit rate on erotic pictures was an artifact of inadequate randomization of their left/right positions.

Because there are distribution assumptions underlying $t$ tests, the significance levels of most of the positive psi results reported in this article were also calculated with nonparametric tests. In this experiment, the hit rates on erotic trials were also analyzed with a binomial test on the overall proportion of hits across all trials and sessions, tested against a null of .5. This is analogous to analyzing a set of coin flips without regard to who or how many are doing the flipping. It is legitimate here because the target was randomly selected on each trial and hence the trials were statistically independent, even within a single session. Across all 100 sessions, the

---

[1] I set 100 as the minimum number of participants/sessions for each of the experiments reported in this article because most effect sizes ($d$) reported in the psi literature range between 0.2 and 0.3. If $d = 0.25$ and $N = 100$, the power to detect an effect significant at .05 by a one-tail, one-sample $t$ test is .80 (Cohen, 1988).

[2] In describing the experiments throughout this article, I have used the plural pronouns "we" and "our" to refer collectively to myself and my research team.

[3] Unless otherwise indicated, all significance levels reported in this article are based on one-tailed tests and $d$ is used as the index of effect size.

53.1% hit rate was also significant by a binomial test ($z$ = 2.30, $p$ = .011).

**Individual differences.** There were no significant sex differences in the present experiment. Over the years, however, the trait of extraversion has been frequently reported as a correlate of psi, with extraverts achieving higher psi scores than introverts. A meta-analysis of 60 independent experiments published between 1945 and 1983, involving several kinds of psi tasks, revealed a small but reliable correlation between extraversion and psi performance ($r$ = .09, $z$ = 4.63, $p$ = .000004; Honorton, Ferrari, & Bem, 1992). The correlation was observed again in a later set of telepathy studies conducted in Honorton's own laboratory, $r$ = .18, $t(216)$ = 2.67, $p$ = .004 (Bem & Honorton, 1994).

The component of extraversion that underlies this correlation appears to be the extravert's susceptibility to boredom and a tendency to seek out stimulation. Eysenck (1966) attributed the positive correlation between extraversion and psi to the fact that extraverts "are more susceptible to monotony . . . [and] respond more favourably to novel stimuli" (p. 59). Sensation seeking is one of the six facets of extraversion on the Revised NEO Personality Inventory (Costa & McCrae, 1992), and Zuckerman's Sensation Seeking Scale (1974), which contains a subscale of Boredom Susceptibility, is significantly correlated with overall extraversion ($r$ = .47, $p < .01$; Farley & Farley, 1967).

To assess stimulus seeking as a correlate of psi performance in our experiments, I constructed a scale comprising the following two statements: "I am easily bored" and "I often enjoy seeing movies I've seen before" (reverse scored). Responses were recorded on 5-point scales that ranged from *Very Untrue* to *Very True* and averaged into a single score ranging from 1 to 5.

In the present experiment, the correlation between stimulus seeking and psi performance was .18 ($p$ = .035). This significant correlation is reflected in the enhanced psi scores of those scoring above the midpoint on the 5-point stimulus-seeking scale: They correctly identified the future position of the picture on 57.6% of the erotic trials, $t(41)$ = 4.57, $p$ = .00002, $d$ = 0.71, exact binomial $p$ = .00008. The difference between their erotic and nonerotic hit rates was itself significant, $t_{\text{diff}}(41)$ = 3.23, $p$ = .001, $d$ = 0.50, with 71% of participants achieving higher hit rates on erotic trials than on nonerotic trials (exact binomial $p$ = .003). Their psi scores on nonerotic trials did not exceed chance, 49.9%, $t(41) = -0.08$, $p$ = .53. Finally, participants low in stimulus seeking did not score significantly above chance on either erotic or nonerotic trials, 49.9%, $t(57) = -0.06$ and 49.9%, $t(57) = -0.13$, respectively.

**But is it precognition? The role of random number generators.** For most psychological experiments, a random number table or the random function built into most programming languages provides an adequate tool for randomly assigning participants to conditions or sequencing stimulus presentations. For both methodological and conceptual reasons, however, psi researchers have paid much closer attention to issues of randomization.

At the methodological level, the problem is that the random functions included in most computer languages are not very good in that they fail one or more of the mathematical tests used to assess the randomness of a sequence of numbers (L'Ecuyer, 2001), such as Marsaglia's (1995) rigorous Diehard Battery of Tests of Randomness. Such random functions are sometimes called pseudo random number generators (PRNGs) because they use a mathematical algorithm to generate each subsequent number from the previous number, and the sequence of numbers is random only in the sense that it satisfies (or should satisfy) certain mathematical tests of randomness. It is not random in the sense of being indeterminate because once the initial starting number (the seed) is set, all future numbers in the sequence are fully determined.

In contrast, a hardware-based or "true" RNG is based on a physical process, such as radioactive decay or diode noise, and the sequence of numbers is indeterminate in the quantum mechanical sense. This does not in itself guarantee that the resulting sequence of numbers can pass all the mathematical tests of randomness, however; some hardware-based RNGs also fail one or more of the tests in the diehard battery (L'Ecuyer, 2001). Both Marsaglia's own PRNG algorithm and the true hardware-based Araneus Alea I RNG used in our experiments pass all his diehard tests.

Note that a random number table is actually a PRNG, even if the sequence of numbers was originally generated by a true RNG. Once the table is printed or stored electronically and an entry point into the table is chosen, the resulting sequence is fully determined, with the entry point being equivalent to the seed number of a computer-based PRNG.

At the conceptual level, the choice of a PRNG or a hardware-based RNG bears on the interpretation of positive findings. In the present context, it bears on my claim that the experiments reported in this article provide evidence for precognition or retroactive influence. In the experiment just reported, for example, there are several possible interpretations of the significant correspondence between the participants' left/right responses and the computer's left/right placements of the erotic target pictures:

1. *Precognition or retroactive influence*: The participant is, in fact, accessing information yet to be determined in the future, implying that the direction of the causal arrow has been reversed.

2. *Clairvoyance/remote viewing*: The participant is accessing already determined information in real time, information that is stored in the computer.

3. *Psychokinesis*: The participant is actually influencing the RNG's placements of the targets.

4. *Artifactual correlation*: The output from the RNG is inadequately randomized, containing patterns that fortuitously match participants' response biases. This produces a spurious correlation between the participant's guesses and the computer's placements of the target picture.

Consider, first, the clairvoyance interpretation. If an algorithm-based PRNG is used for determining the successive left/right positions of the target pictures, then the computer already "knows" the upcoming random number before the participant makes his or her response; in fact, once the initial seed number is generated, the computer implicitly knows the entire sequence of left/right positions. As a result, this information is potentially available to the participant through real-time clairvoyance, permitting us to reject the more extraordinary claim that the direction of the causal arrow has actually been reversed. In contrast, if a true hardware-based

RNG is used for determining the left/right positions, the next number in the sequence is indeterminate until it is actually generated by the quantum physical process embedded in the RNG, thereby ruling out the clairvoyance alternative. This argues for using a true RNG to demonstrate precognition or retroactive influence.

But alas, the use of a true RNG opens up the door to the psychokinesis interpretation: The participant might be influencing the placement of the upcoming target rather than perceiving it, a possibility supported by a body of empirical evidence testing psychokinesis with true RNGs (Radin, 2006, pp. 154–160). Like the clairvoyance interpretation, the psychokinesis interpretation permits us to reject the claim that the direction of the causal arrow has been reversed. Ironically, the psychokinesis alternative can be ruled out by using a PRNG, which is immune to psychokinesis because the sequence of numbers is fully determined and can even be checked after the fact to confirm that its algorithm has not been perturbed. Over the course of our research program—and within the experiment just reported—we have obtained positive results using both PRNGs and a true RNG, arguably leaving precognition/reversed causality the only nonartifactual interpretation that can account for all the positive results.

This still leaves open the artifactual alternative that the output from the RNG is producing inadequately randomized sequences containing patterns that fortuitously match participants' response biases. In the present experiment, this possibility was ruled out by the twin findings that erotic targets were detected significantly more frequently than randomly interspersed nonerotic targets and that the nonerotic targets themselves were not detected significantly more frequently than chance.

Nevertheless, for some of the other experiments reported in this article, it would be useful to have more general assurance that there are not patterns in the left/right placements of the targets that might correlate with response biases of participants. For this purpose, Lise Wallach, professor of psychology at Duke University, suggested that I run a virtual control experiment using random inputs in place of human participants (personal communication, October 10, 2009). In particular, if the human participant is replaced by the same PRNG or RNG that selects the left/right target positions, this maximizes the possibility that any nonrandom patterns in the sequence of left/right target positions will be mirrored by similar patterns in the left/right responses of the virtual participant (the RNG itself), thereby producing an artifactual psi-like result. A null result implies that no such patterns were present.

Accordingly, I ran three experiments that simulated the experiment just reported, one using the internal PRNG of the programming language (REALbasic), one using Marsaglia's PRNG algorithm, and one using the Araneus Alea I hardware-based RNG. Each experiment comprised 100 sessions of 36 trials per session, in which the same PRNG or RNG provided both the left/right response of a virtual participant and the subsequent left/right position of the target. These control experiments all yielded null results: The hit rates were 49.5%, 50.4%, and 49.5% for the internal PRNG, the Marsaglia PRNG, and the Araneus Alea I RNG, respectively (all $t$ values $< 1$). The corresponding correlations (phi) between the virtual participants' input responses and the RNG's target positions were $-.00$, .00, and $-.01$, respectively.

## Experiment 2: Precognitive Avoidance of Negative Stimuli

### Method

One hundred fifty Cornell undergraduates, 107 women and 43 men, were recruited for this experiment through the Psychology Department's automated online sign-up system. Both the recruiting announcement and the opening instructions given to participants upon entering the laboratory, informed them that

> this is an experiment that tests for ESP (Extrasensory Perception). The experiment is run entirely by computer and takes about 15 minutes. . . . On each trial of the experiment you will be shown a picture and its mirror image side by side and asked to indicate which image you like better. The computer will then flash a masked picture on the screen. The way in which this procedure tests for ESP will be explained to you at the end of the session.

Note that the participant's task in this experiment was simply to express a preference between two closely matched pictures on each trial; unlike in traditional precognition experiments (and Experiment 1 reported above), the participant was not faced with an explicit psi challenge.

As in Experiment 1, the participant was seated in front of the computer and asked to respond to the two items on the stimulus seeking scale. This was followed by the same 3-min relaxation period. Then, on each of 36 trials, the participant was shown a low-arousal, affectively neutral picture and its mirror image side by side and asked to press one of two keys on the keyboard to indicate which neutral picture he or she liked better. Using the Araneus Alea I hardware-based RNG, the computer then randomly designated one of the two pictures to be the target. Note that the computer did not determine which of the two neutral pictures would be the target until after the participant had registered his or her preference. Whenever the participant had indicated a preference for the target-to-be, the computer flashed a positively valenced picture on the screen subliminally three times. Whenever the participant had indicated a preference for the nontarget, the computer subliminally flashed a highly arousing, negatively valenced picture. Pictures were again selected primarily from the IAPS set.

The flashed pictures were exposed for 33 ms, followed immediately by a masking stimulus for 167 ms. Time between flashes was 500 ms. A Hubble photograph of the starry sky appeared on the screen for 3,000 ms before the onset of the next trial. A hit was defined as preferring the target-to-be, the picture that avoided the subliminally exposed negative picture. Because participants chose between two images on each trial, the psi hypothesis was that they would prefer the target to the nontarget on significantly more than 50% of the trials.

The pairs of neutral pictures were presented in a fixed order for all sessions, and the RNG randomly determined the left/right position of the two images. For the first 100 sessions, the flashed positive and negative pictures were independently selected and sequenced randomly. For the subsequent 50 sessions, the negative pictures were put into a fixed sequence, ranging from those that had been successfully avoided most frequently during the first 100 sessions to those that had been avoided least frequently. If the participant selected the target, the positive picture was flashed subliminally as before, but the unexposed negative picture was retained for the next trial; if the participant selected the nontarget, the negative picture was flashed and the next positive and negative pictures in the queue were used for the next trial. In other words, no picture was exposed more than once, but a successfully avoided negative picture was retained over trials until it was eventually invoked by the participant and exposed subliminally. The working hypothesis behind this variation in the study was that the psi effect might be stronger if the most successfully avoided negative stimuli were used repeatedly until they were eventually invoked.

## Results and Discussion

The results from the last 50 sessions did not differ significantly from those obtained on the first 100 sessions, so all 150 sessions were combined for analysis.

Unlike in Experiment 1, which had nonerotic trials randomly interspersed among the critical erotic trials, in this experiment every trial constituted a critical trial, making it necessary to confirm that the successive left/right positions of the target were adequately randomized and did not contain patterns that might match participants' response biases. For this reason, I analyzed the data in four ways. The first two analyses were the familiar ones used in Experiment 1: a one-sample $t$ test across participants' hit rates, tested against a null hit rate of 50%, and a nonparametric binomial test on the proportion of hits across all trials and sessions.

The third analysis used an alternative index of psi performance, one that corrects for unequal frequencies of left/right target positions within each session. The output of a session can be represented by the $2 \times 2$ table shown in Figure 1.

The conventional hit rate is defined as hits/trials, or $(A + D)/(A + B + C + D)$. But if there is a bias in the distribution of left/right target positions, if $(A + C) \neq (B + D)$, the hit rate will not necessarily reflect actual psi performance. For example, if the RNG places a majority of the targets on the left and the participant also has a bias favoring pictures on the left, the hit rate will be artifactually inflated. (Experimentally forcing the left/right positions of the target to be equal is not a legitimate corrective for this problem, because it destroys the statistical independence of the trials by using a sampling-without-replacement or "closed deck" procedure.)

An index that automatically corrects for any bias in target position is the phi coefficient, the correlation between the participant's left/right preferences and the RNG's left/right target placements. Phi is computed by the formula $(AD - BC)/[(A + B)(C + D)(A + C)(B + D)]^{1/2}$ and ranges from $-1$ to 1 (Siegel & Castellan, 1988). Phi was computed for each participant, and a $t$ test on phi across participants was calculated. The psi hypothesis is that phi will be significantly greater than 0.

| | Target is on the Left | Target is on the Right |
|---|---|---|
| Participant Prefers Left Picture | A (Hit) | B (Miss) |
| Participant Prefers Right Picture | C (Miss) | D (Hit) |
| | (A + C) | (B + D) |

*Figure 1.* Output of a session represented as a $2 \times 2$ table.

The fourth analysis controlled for any potential overall systematic bias in the RNG by computing an empirical null baseline for each participant rather than assuming it to be 50%. The output of a session in this experiment can be conceptualized as a string of 36 left/right decisions made by the participant compared with the corresponding string of 36 left/right target positions generated by the computer. The number of matches between the two strings is the number of hits obtained. An empirical baseline was computed by running each of the 150 participants through 149 virtual control sessions by comparing his or her decision string with the computer-generated target strings from each of the other 149 (nonself) sessions and calculating the associated hit rates. The mean of those hit rates becomes the empirical baseline for that participant. Because any systematic bias in the RNG will affect the participant's empirical baseline the same way it affects the participant's hit rate, the bias will be subtracted out of the analysis. For descriptive purposes, the difference between the hit rate and the empirical baseline is added to 50% to yield a number that can be directly compared with the more conventional hit rate where 50% is the null.

As Table 2 reveals, the four analyses yielded comparable results, showing significant psi performance across the 150 sessions. Recall, too, that the RNG used in this experiment was tested in the simulation, described above in the discussion of Experiment 1, and was shown to be free of nonrandom patterns that might correlate with participants' responses biases.

**Stimulus seeking.** In the present experiment, the correlation between stimulus seeking and psi performance was .17 ($p = .02$). Table 3 reveals that the subsample of high stimulus seekers achieved an effect size more than twice as large as that of the full sample. In contrast, the hit rate of low stimulus seekers did not depart significantly from chance: 50.7%–50.8%, $t < 1$, $p > .18$, and $d < 0.10$ in each of the four analyses.

# Retroactive Priming

## Experiment 3: Retroactive Priming I

In recent years, priming experiments have become a staple of cognitive and cognitive social psychology (Bargh & Ferguson, 2000; Fazio, 2001; Klauer & Musch, 2003). In a typical affective priming experiment, participants are asked to judge as quickly as they can whether a picture is pleasant or unpleasant, and their response time is measured. Just before the picture appears, a positive or negative word (e.g., *beautiful*, *ugly*) is flashed briefly on the screen; this word is called the prime. Individuals typically respond more quickly when the valences of the prime and the picture are congruent (both are positive or both are negative) than when they are incongruent. In our retroactive version of the procedure, the prime appeared after rather than before participants made their judgments of the pictures.

Because slower responding on congruent trials than on incongruent trials—called a contrast effect—has also been observed in some priming experiments (Hermans, Spruyt, De Houwer, & Eelen, 2003; Klauer, Teige-Mocigemba, & Spruyt, 2009), we also ran a standard nonretroactive priming procedure in each session to ensure that our protocol would produce the usual (noncontrast) priming effect (see also de Boer & Bierman, 2006). Because this turned out to be the case, the psi hypothesis was that the retroactive

Table 2
*Four Analyses of Precognitive Avoidance of Negative Stimuli*

| Hit rate % across participants | Binomial test across trials | Phi coefficient across participants | Hit rate % using an empirical baseline |
|---|---|---|---|
| 51.7%<br>$t(149) = 2.39$<br>$p = .009$<br>$d = 0.20$ | 2,790/5,400 = 51.7%<br>$z = 2.44$<br>$p = .007$<br>$r = .20$ | .034<br>$t(149) = 2.46$<br>$p = .008$<br>$d = 0.20$ | 51.7%<br>$t(149) = 2.37$<br>$p = .010$<br>$d = 0.19$ |

procedure would also produce faster responding on congruent trials than on incongruent trials.

## Method

One hundred Cornell undergraduates, 69 women and 31 men, participated in a 15- to 20-min experiment. They were shown a picture on each of 64 trials and were asked to press one of two keys on the keyboard as quickly as they could to indicate whether the picture was pleasant or unpleasant. The participant's response time in making this judgment was the dependent variable, and the difference in mean response times between incongruent and congruent trials is the index of a priming effect, with positive differences denoting faster responding on congruent trials.

The first 32 trials constituted the retroactive priming procedure, and participants were told that a word would be flashed on the screen just after they made their judgment of the picture. The remaining 32 trials constituted the standard forward priming procedure, and participants were told that "from this point on, the flashed word will appear before rather than after you have made your response." Prior to beginning the actual experimental procedure, participants responded to the two items on the stimulus seeking scale and then had the same 3-min relaxation period described in the previous experiments.

The pictures were again drawn from the IAPS set and were randomly assigned to the forward and retroactive sections of the protocol, with the restriction that an equal number of positive and negative pictures appear in each section. The same 16 positive and 16 negative prime words appeared in both sections, and a prime was randomly selected on each trial before the picture was presented (in the forward priming procedure) or after the participant had responded to the picture (in the retroactive priming procedure). As a result, congruent trials and incongruent trials were randomly sequenced and did not necessarily occur in equal numbers. This made it virtually impossible for participants to anticipate the type of trial coming up by knowing the types of trials that had already occurred. In this experiment, randomizing was implemented by Marsaglia's PRNG algorithm.

Figure 2 displays the time sequence of events for the forward priming and retroactive priming trials, respectively. In both procedures, there was a 2,000-ms interval between trials during which a Hubble photograph of the starry sky appeared on the screen.

## Results and Discussion

Several methods for analyzing response-time data from priming experiments have evolved over the years (Ratcliff, 1993). First, trials on which a participant makes an error in judging the picture to be pleasant or unpleasant are excluded from the analysis. In the present experiment, the median number of errors was three out of 64 trials, and the data from three participants were discarded because they made errors on 16 (25%) or more of the trials, reducing the number of participants to 97. Second, because response-time data are positively skewed, each response time (RT) is usually transformed prior to analysis using either an inverse transformation (1/RT) or a log transformation (log RT). Finally, trials yielding very short or very long response times are considered to be spurious outliers and are excluded from the analysis. Ratcliff (1993) suggested using more than one cutoff criterion to ensure "that an effect is significant over some range of nonextreme cutoffs" (p. 519). Accordingly, Table 4 presents four analyses, using both data transformations and two different cutoff criteria for long response times, 1,500 ms and 2,500 ms. The first criterion excludes 3.1% of the trials; the second excludes 0.5% of the trials.

As shown in the table, the standard forward priming procedure produced the usual result. For example, with a 1,500-ms cutoff criterion and the inverse transformation, participants were 23.6 ms faster on congruent trials than on incongruent trials, $t(96) = 4.91$, $p < .00001$, $d = 0.45$. The retroactive procedure also yielded the predicted psi effect: With a 1,500-ms cutoff criterion and the inverse transformation, participants were 15.0 ms faster on congruent trials than on incongruent trials, $t(96) = 2.55$, $p = .006$, $d = 0.25$. The results were consistent across the range defined by the two cutoff criteria and under both data transformations.

To provide an analysis that avoids distribution assumptions, Table 4 also displays the percentage of participants who had

Table 3
*Four Analyses of Precognitive Avoidance of Negative Stimuli for Participants High in Stimulus Seeking*

| Hit rate % across participants | Binomial test across trials | Phi coefficient across participants | Hit rate % using an empirical baseline |
|---|---|---|---|
| 53.5%<br>$t(49) = 3.07$<br>$p = .002$<br>$d = 0.43$ | 963/1,800 = 53.5%<br>$z = 2.95$<br>$p = .002$<br>$r = .42$ | .079<br>$t(49) = 3.57$<br>$p = .0004$<br>$d = 0.50$ | 53.6%<br>$t(49) = 2.99$<br>$p = .002$<br>$d = 0.39$ |

**Forward Priming Trial**

| Stimulus | Fixation Point | Prime | Blank | Picture | Hubble Photograph |
|---|---|---|---|---|---|
| **Time (ms)** | 1000 | 150 | 150 | Response Time | 2000 |

**Retroactive Priming Trial**

| Stimulus | Fixation Point | Picture | Blank | Prime | Blank | Hubble Photograph |
|---|---|---|---|---|---|---|
| **Time (ms)** | 1000 | Response Time | 300 | 500 | 1000 | 2000 |

*Figure 2.* Timing sequences for events in the retroactive priming procedure.

positive priming scores in each condition, evaluated by an exact binomial test. For example, with a 1,500-ms outlier cutoff criterion, 64.9% of participants produced a positive forward priming effect ($p$ = .002), and 60.8% produced a positive retroactive priming effect ($p$ = .021). There was no significant correlation between stimulus seeking and either priming effect ($r = -.02$ and $-.05$ for the forward and retroactive priming effects, respectively).

## Experiment 4: Retroactive Priming II

### Method

The experiment just described was replicated on an additional 100 participants (57 female and 43 male Cornell undergraduates) with one major change and two minor timing changes. In the original study, the prime paired with each picture was randomly selected on each trial from the list of unused positive and negative primes. In the current replication, one fixed positive prime and one fixed negative prime were assigned to each picture prior to the experiment. These were selected to be semantically relevant to the picture; for example, a picture of a basket of fruit was paired with the positive prime of *luscious* and the negative prime of *bitter,* and a picture of a menacing pit bull was paired with the positive prime of *friendly* and the negative prime of *threatening*. The computer then randomly selected either the positive or the negative prime before the picture was presented (in the forward priming procedure) or after the participant had responded to the picture (in the retroactive priming procedure). In contrast to the randomizing procedure used in the previous priming experiment, this procedure provides a genuine sampling-with-replacement or "open deck" procedure for determining whether a trial will be congruent or incongruent. As a result, there is no (non-psi) way for a participant to anticipate the kind of trial coming up next. All randomizing was again implemented by Marsaglia's PRNG algorithm. In addition, the duration of the fixation point was increased from 1,000 ms to 1,500 ms, and the time between trials (during which the Hubble photograph appeared on the screen) was decreased from 2,000 ms to 1,500 ms.

### Results and Discussion

The data from one participant were discarded because he made errors on more than 16 (25%) of the trials, reducing the number of participants to 99. As seen in Table 5, the results were virtually identical to those obtained in the original experiment. Once again, the standard forward priming procedure produced the usual result. For example, with a 1,500-ms cutoff criterion and the inverse (1/RT) transformation, participants were 27.4 ms faster on congruent trials than on incongruent trials, $t(98) = 4.85$, $p < .00001$, $d = 0.44$. The retroactive procedure also yielded the predicted psi effect again: With a 1,500-ms cutoff criterion and the inverse transformation, participants were 16.5 ms faster on congruent trials than on incongruent trials, $t(98) = 2.03$, $p = .023$, $d = 0.20$. As in Experiment 3, the results were consistent across the different analyses, and the nonparametric exact binomial analyses confirmed both the forward and retroactive priming effects. There was again no correlation between stimulus seeking and either priming effect ($r = -.06$ and $-.07$ for the forward and retroactive priming effects, respectively).

Table 4
*Priming as a Function of Two Outlier Cutoff Criteria and Two Data Transformations*

| Outlier cutoff criterion[a] | >1,500 ms (3.1% of trials excluded)[a] | | >2,500 ms (0.5% of trials excluded)[a] | |
|---|---|---|---|---|
| Transformation | 1/RT | log RT | 1/RT | log RT |
| Forward priming (ms)[b] | 23.6 | | 20.8 | |
| $t(96)$ | 4.91 | 4.47 | 4.71 | 4.00 |
| $p$ | <.00001 | .00001 | <.00001 | .00006 |
| $d$ | 0.45 | 0.42 | 0.43 | 0.38 |
| % with priming > 0 | 64.9 | | 63.9 | |
| $p$ (exact binomial) | .002 | | .004 | |
| Retroactive priming (ms)[b] | 15.0 | | 15.3 | |
| $t(96)$ | 2.55 | 2.49 | 2.57 | 2.42 |
| $p$ | .006 | .007 | .006 | .009 |
| $d$ | 0.25 | 0.25 | 0.25 | 0.24 |
| % with priming > 0 | 60.8 | | 56.7 | |
| $p$ (exact binomial) | .021 | | .111 | |

*Note.* RT = response time.
[a] Also excluded are four trials (0.06%) with response times <250 ms. [b] Incongruent trial RTs minus congruent trials RTs. Calculated prior to applying transformations. Transformations applied only for tests of statistical significance and effect sizes.

Table 5
*Priming as a Function of Two Outlier Cutoff Criteria and Two Data Transformations*

| Outlier cutoff criterion | >1,500 ms (4.7% of trials excluded)[a] | | | >2,500 ms (0.7% of trials excluded)[a] | | |
|---|---|---|---|---|---|---|
| Transformation | 1/RT | | log RT | 1/RT | | log RT |
| Forward priming (ms)[b] | | 27.4 | | | 31.4 | |
| $t(98)$ | 4.85 | | 4.28 | 4.71 | | 3.98 |
| $p$ | <.00001 | | .00002 | <.00001 | | .00007 |
| $d$ | 0.44 | | 0.40 | 0.43 | | 0.37 |
| % with priming > 0 | | 59.6 | | | 59.6 | |
| $p$ (exact binomial) | | .035 | | | .035 | |
| Retroactive priming (ms)[b] | | 16.5 | | | 23.9 | |
| $t(98)$ | 2.03 | | 2.23 | 2.25 | | 2.43 |
| $p$ | .023 | | .014 | .013 | | .008 |
| $d$ | 0.20 | | 0.22 | 0.22 | | 0.24 |
| % with priming > 0 | | 58.6 | | | 61.6 | |
| $p$ (exact binomial) | | .054 | | | .013 | |

*Note.* RT = response time.
[a] Also excluded are seven trials (0.11%) with response times <250 ms. [b] Incongruent trial RTs minus congruent trial RTs. Calculated prior to applying transformations. Transformations applied only for tests of statistical significance and effect sizes.

## Retroactive Habituation and Induction of Boredom

When individuals are initially exposed to a startling or emotionally arousing stimulus, they typically have a strong physiological response to it. Upon repeated exposures, the arousal diminishes, a process known as *habituation*. If the stimulus is initially very unpleasant (e.g., frightening or disgusting), the stimulus becomes more neutral or less negatively arousing; if the stimulus is initially very pleasant, it becomes more neutral or less positively arousing. This was demonstrated in an experiment in which participants who had been subliminally exposed to extremely negative and extremely positive words subsequently rated those words as less extreme than matched control words to which they had not been exposed: Negative words were rated less negatively and positive words were rated less positively than the control words (Dijksterhuis & Smith, 2002).

As Dijksterhuis and Smith noted, the procedure they used in their experiment is similar to that used to demonstrate the well-known mere exposure effect: Across a wide range of contexts, the more frequently humans or other animals are exposed to a particular stimulus, the more they come to like it. This effect has been known for over a century, but it was the publication of Zajonc's (1968) monograph, "Attitudinal Effects of Mere Exposure," that spurred its intensive empirical investigation. In 1989, Bornstein was able to publish a meta-analysis of 208 mere exposure experiments. He reported a mean effect size ($r$) of .26, with a combined $z$ of 20.80 ($p = 2.2 \times 10^{-96}$). Curiously, most mere exposure experiments use low-arousal, affectively neutral stimuli such as nonsense words or polygons; none of the studies reviewed used stimuli that were highly arousing or strongly valenced, either negatively or positively. As a result, the complementary hypothesis that highly arousing positive stimuli might become less liked after repeated exposures had never been tested in a traditional mere exposure experiment.

Among the variables that affect the mere exposure effect are the length of each exposure and the number of exposures. The meta-analysis produced an unequivocal answer for the optimal length of each exposure: The shorter the exposure, the stronger the mere exposure effect, with subliminal exposures being the most effective. This is interpreted as showing that the effect works at an unconscious level and that conscious cognitive processes actually interfere with it. As noted above, subliminal exposures were used in the Dijksterhuis–Smith habituation experiment.

The optimal number of exposures is more difficult to specify because it depends upon several variables, including the complexity of the stimuli and whether different kinds of stimuli are used in the same session (Bornstein, 1989; Bornstein, Kale, & Cornell, 1990). Moreover, the function relating liking to the number of exposures has an inverted-*U* shape: As the number of exposures increases in a mere exposure experiment, liking increases, plateaus, and eventually begins to decline. To account for this, Bornstein proposed a two-process model in which boredom increasingly competes with habituation as the number of exposures increases. Because boredom causes a stimulus to be less liked, the liking curve begins to level off and then turn downward as boredom overtakes habituation. As a result, it is not possible to specify a priori how many exposures would be optimal in any particular experiment. In this experiment, we varied the number of exposures across sessions.[4]

## Experiment 5: Retroactive Habituation I

One hundred Cornell undergraduates, 63 women and 37 men, were recruited through the Psychology Department's automated online sign-up system to serve as participants in a "20–25 minute study of visual imagery that tests for ESP." They either received one point of experimental credit in a psychology course offering that option or were paid $5 for their participation.

[4] This experiment was our first psi study and served as a pilot for the basic procedures adopted in all the other studies reported in this article. When it was conducted, I had not yet introduced the hardware-based random number generator or the stimulus seeking scale. Preliminary results were reported at the 2003 convention of the Parapsychological Convention in Vancouver, Canada (Bem, 2003); subsequent results and analyses have revised some of the conclusions presented there.

Upon entering the laboratory, the participant was told,

> In this experiment, we are interested in measuring emotional reactions to a wide variety of visual images in a procedure that tests for ESP (Extrasensory Perception). The experiment is run completely by a computer and takes about 20–25 minutes.
>
> Each trial of the experiment involves a pair of pictures. First you will be shown the two pictures side by side and asked to indicate which one you like better. You will then be asked to watch passively as those pictures are flashed rapidly on the screen. The way in which this procedure tests for ESP will be explained at the end of the session.
>
> Most of the pictures range from very pleasant to mildly unpleasant, but in order to investigate a wide range of emotional content, some of the pictures contain very unpleasant images (e.g., snakes and bodily injuries).

The participant then signed a consent form that repeated the warning about the nature of the pictures. (The same warning also appeared in the online recruiting page.) Next, the experimenter seated the participant in front of the computer and withdrew from the cubicle.

## Method

This experiment adapted a binary-choice version of a mere exposure experiment (Kunst-Wilson & Zajonc, 1980) and, in effect, ran it backwards. In a standard version of the experiment, the following sequence of events would occur:

1. The participant is repeatedly exposed subliminally to a picture. This picture is called the habituation target.

2. The participant is then shown two pictures side by side and asked to indicate which one he or she likes better. One of the pictures is the target; the other is a closely matched picture that the participant has not seen before. If the participant prefers the target, the trial is scored as a hit. The hit rate expected by chance is thus 50%.

The retroactive version of this protocol simply reverses Steps 1 and 2: On each trial, the participant is first shown a pair of matched pictures on the computer screen and asked to indicate which picture he or she prefers. The computer then randomly selects one of the two pictures to serve as the habituation target and displays it subliminally several times. This first retroactive habituation experiment comprised trials using either strongly arousing negative picture pairs or neutral control picture pairs; positively arousing (i.e., erotic) picture pairs were not introduced until Experiment 6, reported below. The retroactive habituation hypothesis was that on trials with negative picture pairs, participants would prefer the target to the nontarget on more than 50% of the trials.

As in the experiments reported above, most of the pictures for this study were selected from the IAPS set. The pictures in each pair were matched not only for content but for valence and arousal using the ratings supplied with the IAPS set.

**Procedure.** The computer program administered the procedure as outlined above. After the usual relaxation procedure, the participant was shown two matched pictures side by side on the screen on each of 48 trials and asked to indicate which of the pair he or she liked better by pressing one of two keys on the keyboard. Each pair consisted of two high-arousal negative pictures or two low-arousal, affectively neutral pictures; the left/right position of the two pictures was randomly determined. The computer then randomly selected one of the two pictures to be the target and flashed it randomly on the left or the right side of the screen for 17 ms, followed immediately by a masking stimulus that remained on the screen for 33 ms. There was a 1,000-ms blank interval between exposures. In this experiment, randomizing was implemented through the programming language's internal random function.

The number of exposures varied, assuming the values of 4, 6, 8, or 10 across this experiment and its replication. Note that participants were not aware that only one of the two pictures was flashed on each trial, because the target was exposed subliminally and the instructions implied that both pictures of the pair would be flashed.

## Results and Discussion

The retroactive habituation hypothesis was supported. On trials with negative picture pairs, participants preferred the target significantly more frequently than the nontarget, 53.1%, $t(99) = 2.23$, $p = .014$, $d = 0.22$. This is confirmed by a binomial test on the proportion of hits across all trials with negative picture pairs, .531, $z = 2.09$, $p = .018$. Women achieved a significant hit rate on the negative pictures, 53.6%, $t(62) = 2.25$, $p = .014$, $d = 0.28$, but men did not, 52.4%, $t(36) = 0.89$, $p = .19$, $d = 0.15$. This sex difference is not statistically significant, $t_{\text{diff}}(62) = 0.39$, $p = .70$, two-tailed, but it did prompt us to introduce different pictures for men and women in the replication (Experiment 6) reported below. Participants' hit rates on the control trials with neutral picture pairs did not differ from chance expectation, 49.4%, $t(99) = -0.73$, $p = .46$, two-tailed.

# Experiment 6: Retroactive Habituation II

One hundred fifty Cornell undergraduates, 87 women and 63 men, participated in this replication and extension of Experiment 5. Most important, this replication added trials with erotic picture pairs. The retroactive habituation hypothesis for these trials was the opposite of that for negative trials: Participants would prefer the target picture—the one to be repeatedly exposed—on *less* than 50% of the erotic trials.

There were other two other changes. First, on the basis of the preference data obtained in Experiment 5, we were able to better equate the "popularity" of the two pictures within each pair (i.e., the frequency with which they had been preferred in the previous sessions). In a few cases, new pictures from the IAPS set were substituted. Second, we decided to use sets of negative and erotic pictures that were different for men and women. As noted above, women showed a significant psi effect on the negative trials in Experiment 5, but men did not. Because the psi literature does not reveal any systematic sex differences in psi ability, it seemed possible that the men were simply less aroused than the women by the negative pictures. The ratings supplied with the IAPS pictures revealed that male raters rated every one of the negative pictures in the set as less negative and less arousing than did female raters. Also, an fMRI study using IAPS pictures found that men had significantly fewer brain regions than women where activation correlated with concurrent ratings of their emotional experience

(Canli, Desmond, Zhao, & Gabrieli, 2002). So, for this replication, we supplemented the IAPS pictures for men with stronger and more explicit negative and erotic images obtained from Internet sites. We also provided two additional sets of erotic pictures, so that men could choose the option of seeing male–male erotic images and women could choose the option of seeing female–female erotic images.

## Results and Discussion

Both retroactive habituation hypothesis were supported. On trials with negative picture pairs, participants preferred the target significantly more frequently than the nontarget, 51.8%, $t(149) = 1.80$, $p = .037$, $d = 0.15$, binomial $z = 1.74$, $p = .041$, thereby providing a successful replication of Experiment 5. On trials with erotic picture pairs, participants preferred the target significantly less frequently than the nontarget, 48.2%, $t(149) = -1.77$, $p = .039$, $d = 0.14$, binomial $z = -1.74$, $p = .041$.

An overall psi score that combines the two complementary effects can be computed by subtracting the erotic hit rate from the negative hit rate: This yields a difference of 3.76%, $t_{\mathrm{diff}}(149) = 2.41$, $p = .009$, $d = 0.20$. Because negative and erotic trials were randomly interspersed in the trial sequence, this significant difference also serves as evidence against the possibility that the significant hit rates were an artifact of inadequate randomization of left/right target positions. On the neutral control trials, participants scored at chance level: 49.3%, $t(149) = -0.66$, $p = .51$, two-tailed. Hit rates were not significantly correlated with the number of exposures of the target picture across the two experiments, and there were no significant sex differences on any of the measures.

**Erotic stimulus seeking.** Although I had not yet introduced the stimulus seeking scale into our research program when the two retroactive habituation experiments were conducted, I did define a measure of erotic stimulus seeking for this replication by converting two items from Zuckerman's Sensation Seeking Scale (1974) into true/false statements: "I enjoy watching many of the erotic scenes in movies" and "I prefer to date people who are physically exciting rather than people who share my values." A participant's score was simply the number of items endorsed, and participants who endorsed both statements were defined as erotic stimulus seekers. These two items were administered to 100 of the 150 participants in this replication prior to the relaxation period and experimental trials. There was a highly significant positive correlation between erotic stimulus seeking and psi performance (lower hit rates), $r = .24$, $p = .008$, with erotic stimulus seekers showing a very strong retroactive habituation effect on the erotic trials, 43.1%, $t(31) = -3.20$, $p = .002$, $d = 0.57$, exact binomial $p = .002$. The hit rate of those who were not erotic stimulus seekers did not differ from chance: 51.3%, $t(67) = 0.78$, $p = .78$, $d = -0.09$.

**Control trials.** As reported above, overall hit rates on control trials did not differ from chance in either of the two experiments. This might appear inconsistent with the results of mere exposure studies in which neutral, nonarousing stimuli typically show increased liking. Typically but not inevitably. Experiments reported by Bornstein et al. (1990) demonstrated that increased liking for such stimuli "are only produced when subjects are not currently exposed to stimuli that are more interesting or complex. In other words, robust exposure effects for simple stimuli depend on their *not* being presented along with more interesting stimuli on a within-subjects basis" (p. 798). In our experiment, the control stimuli were arguably less interesting than either the negative or erotic stimuli presented concurrently in the same session, so our findings on control trials are not inconsistent with the findings from mere exposure experiments. This same contrast effect may also explain why participants in Experiment 1 were able to significantly detect the position of future erotic pictures but not that of nonerotic negative, neutral, or positive pictures.

The control trials also yielded a significant serendipitous finding. The hit rate on control trials was at chance for exposure frequencies of 4, 6, and 8. On sessions with 10 exposures, however, it fell to 46.8%, $t(39) = -2.12$, two-tailed $p = .04$, binomial $z = -2.17$, two-tailed $p = .03$. This effect can be interpreted as the retroactive induction of boredom. As with a too frequent TV commercial, the many repeated exposures retroactively rendered the neutral target picture boring, or even aversive, and hence less attractive than its matched nontarget. In other words, 10 exposures was the point at which the inverted-*U*-shaped function between liking and exposure frequency began its downturn for neutral control stimuli. This is presumably the same effect as the decreased liking observed on erotic trials: Repeated exposures of an erotic target diminish its erotic potency and render it boring relative to its matched nontarget. This apparently occurs at lower exposure frequencies for erotic stimuli than it does for neutral control stimuli.

# External Replications

There have been two published replications of the retroactive habituation effect. Savva, Child, and Smith (2004) conducted a conceptual replication in which all the negative pictures were images of spiders and the neutral pictures were images of landscapes. No erotic pictures were used. The participants were 25 spider-phobic and 25 non-spider-phobic individuals. The spider-phobic participants had a hit rate significantly above chance on the spider trials, 53.7%, $t(24) = 1.70$, $p = .05$, $d = 0.34$, which was also significantly higher than the 48.2% hit rate on the neutral control trials, $t_{\mathrm{diff}}(24) = 2.48$, $p = .010$, $d = 0.50$. The hit rate of the non-spider-phobic participants was at chance on both spider and control trials.

Parker and Sjödén (2010) also conducted a replication using only negative and neutral pictures, but their participants went through both the retroactive habituation procedure and an adaptation of Dijksterhuis and Smith's (2002) original habituation procedure. The same stimulus pictures were used in both. In the regular habituation procedure, participants were subliminally exposed six times to half of the pictures and were then asked to rate the valence of both the exposed and the nonexposed pictures.

Overall, there was not a significant retroactive habituation effect on the negative pictures, 51.0%, $t(49) = 0.51$, $p = .31$, but there was a highly significant correlation between habituation and retroactive habituation to the negative pictures ($r = .34$, $p = .008$). The 34 participants who showed habituation also showed significant retroactive habituation, 53.9%, $t(33) = 1.93$, $p = .031$, $d = 0.33$; the 16 participants who failed to show habituation also failed to show retroactive habituation, 44.8%, $t(15) = -1.32$, $p = .90$.

## Experiment 7: Retroactive Induction of Boredom

The serendipitous finding that sessions with 10 exposures produced a hit rate significantly below 50% on neutral control trials suggested that it might be possible to design an experiment specifically designed to produce retroactive induction of boredom on nonarousing neutral stimuli as the central phenomenon. This would be desirable for at least two reasons. First, there are large age, cultural, and individual differences in reactions to the kinds of negative and erotic pictures used in the retroactive habituation experiment, making successful replication across different populations less likely. Second, some psi researchers (or their institutional review boards) have been reluctant to conduct experiments in which participants are exposed to these kinds of pictures. This experiment tests directly for the retroactive induction of boredom on neutral stimuli.

In a mere exposure experiment that included a measure of "boredom proneness," Bornstein et al. (1990) report that only participants who were not prone to boredom showed a significant mere exposure effect (i.e., increased liking for a frequently exposed stimulus). This suggests that boredom dominated habituation for boredom prone participants. It should be apparent that our two-item stimulus seeking scale ("I am easily bored" and "I often enjoy seeing movies I've seen before"; reverse scored) could equally well be conceptualized as an index of boredom proneness. For this reason, I first introduced it into the current experiment, with the corresponding hypothesis that those high in stimulus seeking (high in boredom proneness) would show significantly decreased liking for the target.

### Method

Two hundred Cornell undergraduates, 140 women and 60 men, were recruited for this experiment. From the IAPS we selected sets of matched picture pairs that ranged from mildly negative to positive; no strongly negative or erotic pictures were included.

After the program had administered the two stimulus-seeking items and provided the 3-min relaxation period, the protocol was essentially the same as that used in the retroactive habituation experiments. On each of 24 trials, the participant was shown two matched pictures and asked to click the mouse on the picture he or she preferred. The computer then randomly selected one of the two pictures to serve as the target and flashed it on the screen 10 times. Unlike in the retroactive habituation experiments, however, the exposures were supraliminal (750-ms duration followed by a blank screen for 250 ms) and were enlarged to fill the entire screen. It was my (wrongheaded) hunch that supraliminal exposures would be more likely to produce boredom after 10 exposures than would the subliminal exposures successfully used in our original retroactive habituation experiments. Also, because 10 supraliminal exposures on each trial take up a lot of time, we had to limit the number of trials to 24 to avoid rendering the entire experiment boring, not just the frequently exposed stimulus pictures.

The random sequencing of the picture pairs, the left/right placement of the two pictures, and the selection of the target were all implemented with Marsaglia's PRNG algorithm.

### Results and Discussion

Across all 200 sessions, the hit rate was in the predicted direction but not significantly different from chance, 49.1%, $t(199) = -1.31$, $p = .096$, $d = 0.09$. (I now wish I had simply continued to use subliminal exposures.) Nevertheless, stimulus seeking was again positively correlated with psi performance (lower hit rates), $r = .16$, $p = .011$. Participants high in stimulus seeking obtained a hit rate significantly below chance, 47.9%, $t(95) = -2.11$, $p = .019$, $d = 0.22$, binomial $z = -1.94$, $p = .026$, whereas the remaining participants did not, 50.1%, $t(103) = 0.17$, $p = .43$.

As in Experiment 2 on the precognitive avoidance of negative stimuli, there were no control trials randomly interspersed among the critical trials in this experiment, making it necessary again to confirm that the successive left/right positions of the targets were adequately randomized and did not contain patterns that might match participants' response biases. Accordingly, we analyzed the data on the stimulus-seeking subsample in the same four ways as in Experiment 2. The first two analyses were the familiar ones used previously: a one-sample *t* test across participants' hit rates, tested against a null hit rate of 50%, and a nonparametric binomial test on the proportion of hits across all trials and sessions for the 96 participants high in stimulus seeking.

The third analysis used the phi coefficient, the correlation between the participant's left/right preference and the PRNG's left/right target placements, as the index of psi performance. And finally, the fourth analysis controlled for any potential systematic bias in the PRNG by computing an empirical null baseline for each participant, as described in Experiment 2.

As Table 6 reveals, all four analyses yielded comparable results, showing a significant retroactive boredom induction for participants high in stimulus seeking. It will also be recalled from the discussion of Experiment 1 that Marsaglia's PRNG algorithm was shown in the simulation with random inputs to be free of nonrandom patterns that might correlate with participants' responses biases. The hit rate of participants low in stimulus seeking did not depart significantly from chance in any of the four analyses, with hit rates ranging from 50.1% to 50.2% (all *t* values $< 1$, all *p*s $>$ .40, and all *d*s $< 0.02$).

Table 6
*Four Analyses of Retroactive Induction of Boredom for Participants High in Stimulus Seeking*

| Hit rate % across participants | Binomial test across all trials | Phi coefficient across participants | Hit rate % using an empirical baseline |
|---|---|---|---|
| 47.9%<br>$t(95) = -2.11$<br>$p = .019$<br>$d = 0.22$ | 1,105/2,304 = 47.9%<br>$z = -1.94$<br>$p = .026$<br>$r = .20$ | −.025<br>$t(95) = -2.14$<br>$p = .017$<br>$d = 0.22$ | 48.0%<br>$t(95) = -2.10$<br>$p = .019$<br>$d = 0.21$ |

## Retroactive Facilitation of Recall

In Lewis Carroll's *Through the Looking Glass*, the White Queen explains to Alice that the citizens of her country have precognitive ability; or, as she puts it, "memory works both ways" in her land and she herself remembers best "things that happened the week after next." When Alice says that "I'm sure mine only works one way . . . I can't remember things before they happen," the Queen disparagingly remarks, "It's a poor sort of memory that only works backwards" (Carroll, 2006, p. 164).

## Experiment 8: Retroactive Facilitation of Recall I

Inspired by the White Queen's claim, the current experiment tested the hypothesis that memory can "work both ways" by testing whether rehearsing a set of words makes them easier to recall—even if the rehearsal takes place after the recall test is given. Participants were first shown a set of words and given a free recall test of those words. They were then given a set of practice exercises on a randomly selected subset of those words. The psi hypothesis was that the practice exercises would retroactively facilitate the recall of those words, and, hence, participants would recall more of the to-be-practiced words than the unpracticed words.

### Method

One hundred Cornell undergraduates, 64 women and 36 men, participated. Upon entering the laboratory, the participant was told, "This experiment tests for ESP by administering several tasks involving common everyday words. The experiment is run completely by computer and takes about 20 minutes. The program will give you specific instructions as you go. At the end of the session, I will explain to you how this procedure tests for ESP." Note that the participant was not told in advance about the recall test.

Participants were then seated in front of the computer. After they had responded to the two stimulus-seeking items and gone through the 3-min relaxation procedure described in previous experiments, they were shown 48 common nouns one at a time for 3 s each. These words were drawn from four categories (foods, animals, occupations, and clothes) and were presented in the same fixed order for all participants. They were asked to visualize the referent of each word as it appeared on the screen (e.g., if the word was *tree*, they were to visualize a tree). They were then given a (surprise) free recall test in which they were asked to type all the words they could recall in any order.

After the participant completed the recall test, the computer randomly selected six words from each of the four categories to serve as practice words, with the remaining 24 words serving as no-practice control words. The 24 practice words all appeared together in a randomized list on the screen. The participant was informed that the words were drawn from four categories and asked to click on the six food words in the list (which turned red when clicked) and then to retype those words into six empty slots on the screen. The list was rescrambled, and the same task was repeated for each of the other three categories of words. In all, the participant was required to scan the list of practice words four times, to click on the six words in each category, and to type out each of the 24 words.

### Results and Discussion

Unlike in a traditional experiment in which all participants contribute the same fixed number of trials, in the recall test each word the participant recalls constitutes a trial and is scored as either a practice word or a control word. Accordingly, the measure of psi in this study is a weighted differential recall (DR) score, defined as the number of practice words recalled minus the number of control words recalled (P – C) multiplied by the participant's overall recall score (P + C). This gives more weight to participants who recalled more words ("contributed more trials") and is conceptually analogous to the practice of weighting studies by their sample sizes in a meta-analysis. For descriptive purposes, the DR score is expressed as a percentage of the maximum possible DR score (= 576), which would be achieved if a participant recalled all 24 practice words but none of the 24 control words. Thus, $DR\% = [(P - C) \times (P + C)]/576$ and can range from −100% to 100%, with positive DR% scores denoting that more practice words were recalled than control words.

The results show that practicing a set of words after the recall test does, in fact, reach back in time to facilitate the recall of those words: The mean DR% for the total sample was 2.27%, $t(99) = 1.92$, $p = .029$, $d = 0.19$. Once again, stimulus seeking was significantly correlated with psi performance (DR%): $r = .22$, $p = .014$. This correlation is reflected in the strong DR% scores of participants high in stimulus seeking: 6.46%, $t(42) = 3.76$, $p = .0003$, $d = 0.57$. In contrast, those low in stimulus seeking scored at chance level: −0.90%, $t(56) = -.60$, $p = .73$.

**Control sessions.** Although no control group was needed to test the psi hypothesis in this experiment, we ran 25 control sessions in which the computer again randomly selected a 24-word practice set but did not actually administer the practice exercises. These control sessions were interspersed among the experimental sessions, and the experimenter was uninformed as to condition.

The main purpose in the control sessions was to see if the practice exercises retroactively produced increased clustering in the recall test. Clustering is a spontaneous effect observed in free recall studies in which recalling a word from a particular category (e.g., foods) is more likely to be followed by recalling another word from the same category than by recalling a word from a different category (e.g., animals; Bousfield, 1953; Cofer, 1965). The original sequence of the 48 words given prior to the recall test contained no clustering (i.e., no two successive words belonged to the same category). Because participants are alerted by the practice exercises to the fact that the words come from four distinct categories, they might be expected to show more clustering than participants in the control sessions.

A second purpose in the control sessions was to see whether the practice exercises retroactively enhanced overall recall. And finally, the control sessions enabled us to confirm that it is the actual practice that produces the psi effect, not just the existence of a computer-chosen set of practice words that participants might be able to access clairvoyantly.

The results were clear. There was no more clustering in the experimental sessions than in the control sessions: In both, 36% of successive pairs of recalled words came from the same category. Similarly, overall recall was not higher in the experimental sessions: In both conditions, participants recalled 18.4 (38.3%) of the 48 words in the original list. This means that the enhanced recall of practice words came at the expense of diminished recall of control words. This is a well-known phenomenon in memory research and is called retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994). It is embodied in the joke about the professor of ichthyology who complained that whenever she memorized the name of a student, she forgot the name of a fish. As expected, the mean DR% score from control sessions did not differ from zero for the sample as a whole, 0.26%, $t(24) = 0.13$, $p = .45$, or for participants high in stimulus seeking, $-.09\%$, $t(10) = -0.02$, $p = .51$.

## Experiment 9: Retroactive Facilitation of Recall II

### Method

Experiment 9 was a replication of Experiment 8 with one procedural change: A new practice exercise was introduced immediately following the recall test in an attempt to further enhance the recall of the practice words. This exercise duplicated the original presentation of each word that participants saw prior to the recall test, but only the practice words were presented. The instructions were as follows: "You will now be shown 24 of the words you saw earlier, divided into 4 categories: Foods, Animals, Occupations, and Clothing. As you see each word, try to form an image of the thing it refers to (e.g., if the word is *tree*, visualize a tree)."

As in the original presentation prior to the recall test, each word was displayed for 3 s, but this time all the words from each category were presented consecutively. That is, the six food words were displayed first, followed by the six animal words, and so forth. Except for this additional practice exercise, the experiment was otherwise identical to Experiment 8. A total of 50 Cornell undergraduates participated, 34 women and 16 men. In addition, we again ran 25 control sessions.

### Results and Discussion

This modified replication yielded an even stronger psi effect than that in the original experiment: The mean DR% score was 4.21%, $t(49) = 2.96$, $p = .002$, $d = 0.42$, an effect size more than twice the effect size of 0.19 found in the original experiment. In this replication, however, stimulus seeking was no longer significantly correlated with psi performance ($r = -.10$, $p = .25$). It appears that the strong stimulus manipulation that produced the higher effect size also restricted the range of DR% scores sufficiently to squelch the predictive power of the individual-difference measure: The range of DR% scores was 42% smaller in the replication than in the original experiment; the variance was 27% smaller. Both those high and those low in stimulus seeking obtained significant DR% scores, 4.47%, $t(15) = 1.77$, $p = .049$, $d = 0.44$ and 4.09%, $t(33) = 2.34$, $p = .013$, $d = 0.40$, respectively.

As in the original experiment, clustering was no higher in the experimental sessions (39%) than in the control sessions (40%), despite the complete clustering in the presentation sequence of words in the new practice exercise. Similarly, overall recall was again not significantly higher in the experimental sessions (45%) than in the control sessions (44%). And finally, DR% scores from the control sessions did not differ significantly from zero for the sample as a whole or for participants high in stimulus seeking.[5]

## General Discussion

Table 7 summarizes the significance levels and effect sizes obtained in the nine experiments reported in this article. As seen in the table, the mean effect size across all the experiments was .22, and all but the retroactive induction of boredom experiment yielded statistically significant results. Stimulus seeking was significantly correlated with psi performance in five of the experiments (including the induction of boredom experiment), and these correlations are reflected in the enhanced psi performances across those experiments by those high in stimulus seeking: For the stimulus-seeking subsamples, the mean effect size across all experiments in which the stimulus-seeking scale was administered was 0.43.

### Research Strategy

As noted in the introduction, the goal in this research program was to develop well-controlled demonstrations of psi that could be replicated by independent investigators. The research strategy was to design experiments and employ statistical procedures that were as simple, as transparent, and as familiar as possible. First, the experimental procedures were based on simple, well-established psychological effects that would be familiar to most readers of this journal. Second, the primary statistical tool for evaluating the results was the familiar one-sample *t* test across sessions; however, because *t* tests rely on the assumption that the data are normally distributed, they were supplemented by equally simple and familiar nonparametric binomial tests.

There are, of course, more sophisticated statistical techniques available for dealing with distribution issues in hypothesis testing (e.g., Wilcox, 2005), but they do not yet appear to be widely familiar to psychologists and are not yet included in popular statistical computer packages, such as SPSS. I have deliberately not used them for this article. It has been my experience that the use of complex or unfamiliar statistical procedures in the reporting of psi data has the perverse effect of weakening rather than strengthening the typical reader's confidence in the findings. The same is true of complex or unfamiliar experimental procedures.

This is understandable. If one holds low Bayesian a priori probabilities about the existence of psi—as most academic psychologists do—it might actually be more logical from a Bayesian perspective to believe that some unknown flaw or artifact is hiding in the weeds of a complex experimental procedure or an unfamiliar statistical analysis than to believe that genuine psi has been demonstrated. As a consequence, simplicity and familiarity become essential tools of persuasion. I return to the theme of familiarity and belief in psi in the section on physics below.

Finally, I have analyzed and reported alternative indices of psi performance (e.g., both the hit rate and phi in some of the binary choice experiments) and alternative treatments of the data (e.g., different outlier cutoff criteria and data transformations in the priming experiments) to confirm the consistency of the results across these variations. I hope that this will allay the common concern that an investigator may have tried many data analyses but

[5] A small pilot study ($N = 38$) in Sweden, using English words and a different practice procedure, failed to find significant retroactive facilitation of recall (Cardeña, Marcusson-Clavertz, & Wasmuth, 2009).

Table 7
*Psi Performance in All Nine Experiments: Probability Levels (p), Effect Sizes (d), and Correlations (r) With Stimulus Seeking (SS)*

| Phenomenon tested and experiment | *p* full sample | *d* full sample | Correlation with SS | *p* high SS | *d* high SS | *p* low SS | *d* low SS |
|---|---|---|---|---|---|---|---|
| Precognitive approach/avoidance | | | | | | | |
| 1. Detection of Erotic Stimuli | .01 | 0.25 | .18* | .00002 | 0.71 | .524 | −0.01 |
| 2. Avoidance of Negative Stimuli[a] | .009 | 0.20 | .17** | .001 | 0.45 | .215 | 0.08 |
| Retroactive priming | | | | | | | |
| 3. Retroactive Priming I[a] | .007 | 0.26 | −.05 | .148 | 0.17 | .036 | 0.24 |
| 4. Retroactive Priming II[a] | .014 | 0.23 | −.07 | .059 | 0.27 | .035 | 0.23 |
| Retroactive habituation | | | | | | | |
| 5. Retroactive Habituation I | | | | | | | |
| Negative trials[b] | .014 | 0.22 | | | | | |
| 6. Retroactive Habituation II | | | | | | | |
| Negative trials[b] | .037 | 0.15 | | | | | |
| Erotic trials | .039 | 0.14 | 0.24*** | .002 | 0.57 | .219 | −0.09 |
| 7. Retroactive Induction of Boredom[a] | .096 | 0.09 | .16** | .018 | 0.22 | .483 | 0.00 |
| Retroactive facilitation of recall | | | | | | | |
| 8. Facilitation of Recall I | .029 | 0.19 | .22** | .0003 | 0.57 | .525 | −0.08 |
| 9. Facilitation of Recall II | .002 | 0.42 | −.10 | .049 | 0.44 | .013 | 0.40 |
| Mean effect size (*d*) | | 0.22 | | | 0.43 | | 0.10 |

[a] Probabilities and effect sizes in this row are based on the mean of the *t* values across the variations of the data analysis. [b] The Stimulus-Seeking Scale was not administered in this experiment.
* $p < .05$. ** $p < .02$. *** $p < .01$.

reported only the one that "worked"—a variant of the file-drawer problem.

## The File Drawer

Like most social-psychological experiments, the experiments reported here required extensive pilot testing. As all research psychologists know, many procedures are tried and discarded during this process. This raises the question of how much of this pilot exploration should be reported to avoid the file-drawer problem, the selective suppression of negative or null results.

This problem arose most acutely in our two earliest experiments, the retroactive habituation studies, because they required the most extensive pilot testing and served to set the basic parameters and procedures for all the subsequent experiments. I can identify three sets of findings omitted from this report so far that should be mentioned lest they continue to languish in the file drawer.

First, several individual-difference variables that had been reported in the psi literature to predict psi performance were pilot tested in these two experiments, including openness to experience; belief in psi; belief that one has had some psi experiences in everyday life; and practicing a mental discipline such as meditation, yoga, self-hypnosis, or biofeedback. None of them reliably predicted psi performance, even before application of a Bonferroni correction for multiple tests. Second, an individual-difference variable (negative reactivity) that I reported as a correlate of psi in my convention presentation of these experiments (Bem, 2003) failed to emerge as significant in the final overall database.

Finally, as also reported in Bem (2003), I ran a small retroactive habituation experiment that used supraliminal rather than subliminal exposures. It was conducted as a matter of curiosity after the regular (subliminal) experiment and its replication had been successfully completed. It yielded chance findings for both negative and erotic trials. As I warned in the convention presentation, supraliminal exposures fundamentally change the phenomenology of the experiment for participants. First, they become aware that only one of the two original pictures is being repeatedly exposed; second, supraliminal exposures provide explicit trial-by-trial feedback, prompting many participants to become involved in anticipating which picture will appear. This transforms the experiment for them into an explicit ESP challenge and undermines the very rationale for using an implicit response measure of psi in the first place. Unfortunately, I ignored this warning myself when I subsequently used supraliminal exposures in the experiment on the retroactive induction of boredom.

## Issues of Replication

Replication packages are available on request for Macintosh and Windows-based computers to encourage and facilitate replication of the experiments reported here. I suspect that the experiments on retroactive priming and retroactive facilitation of recall will be the easiest to replicate successfully and that the experiments using erotic and negative stimuli will be more difficult to replicate, especially on subject populations markedly different from the North American college-age men and women who participated in our experiments. Their cohort's distinctive experiences with erotic and violent imagery may render their responses to such stimuli unrepresentative of older ages or other cultures.

More general issues of replication are also pertinent here. Although most academic psychologists do not believe in psi, many apparently do believe in miracles when it comes to replication. Tversky and Kahneman (1971) posed the following problem to their colleagues at meetings of the Mathematical Psychology Group and the American Psychological Association: "Suppose you have run an experiment on 20 subjects and have obtained a significant result which confirms your theory ($z = 2.23$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group?" (p. 105).

The median estimate was .85, with nine out of 10 respondents providing an estimate greater than .60. The correct answer is approximately .48. As Rosenthal (1990) has warned: "Given the levels of statistical power at which we normally operate, we have no right to expect the proportion of significant results that we typically do expect, even if in nature there is a very real and very important effect" (p. 16). One implication of this warning is that replication studies with insufficient power are simply not informative when they fail to replicate the original findings.

One of the major obstacles to successful replication in psychology generally is the influence of the experimenter on the results. The sex, age, and demeanor of the experimenter can interact with characteristics of the participants, and expectancies of the experimenter can affect the results in subtle ways (Rosenthal, 1966). Psi research is no exception. In three psi experiments specifically designed to investigate the experimenter effect, a proponent and a skeptic of psi jointly ran a psi experiment, using the same procedures and drawing participants from the same pool (Schlitz, Wiseman, Radin, & Watt, 2005; Wiseman & Schlitz, 1997, 1999). In two of the three experiments, the proponent obtained a significant result, but the skeptic did not.

My approach to the problem of experimenter effects has been to minimize the experimenter's role as much as possible, reducing it to that of greeter and debriefer, and leaving the experimental instructions and other interactions with the participant to the computer program. Moreover, I used several undergraduate experimenters in each experiment and deliberately gave them only informal training. This was to ensure that the experimental protocols were robust enough to overcome differences among experimenters, so that the protocols have a better chance of surviving replications in other laboratories. Whether or not this strategy will be successful remains to be seen.

Finally, the success of replications in psychological research often depends on subtle and unknown factors. For example, Bornstein's (1989) meta-analysis of the well-established mere exposure effect reveals that the effect fails to replicate on simple stimuli if other, more complex stimuli are presented in the same session. It also fails to replicate if too many exposures are used, if the exposure duration is too long, if the interval between exposure and the assessment of liking is too short, or if participants are prone to boredom. As previously noted, the mere exposure effect had not even been tested with strongly valenced stimuli until Dijksterhuis and Smith (2002) conducted their habituation experiment, showing that strong positive stimuli actually reverse the mere exposure effect.

My intent here is not to offer preemptive excuses for future failures to replicate the results reported in this article, but rather to urge that we not impose an unrealistically high standard—or a double standard—on the replication of psi effects.

## The Psychology of Psi

If psi exists, then it is not unreasonable to suppose that it might have been acquired through evolution by conferring survival and reproductive advantage on the species (for a discussion, see Broughton, 1991, pp. 347–352). For example, the ability to anticipate and thereby to avoid danger confers an obvious evolutionary advantage that would be greatly enhanced by the ability to anticipate danger precognitively. It was this reasoning that motivated Experiment 2 on the precognitive avoidance of negative stimuli. Similarly, the possibility of an evolved precognitive ability to anticipate sexual opportunities motivated Experiment 1 on the precognitive detection of erotic stimuli. The presentiment experiments were probably inspired by similar reasoning.

Evolutionary reasoning also played a part in Eysenck's (1966) discussion of psi and extraversion, cited earlier in this article. In particular, he speculated that psi might be a primitive form of perception antedating cortical developments in the course of evolution, and, hence, cortical arousal might suppress psi functioning. Because extraverts have a lower level of cortical arousal than introverts, that provides another reason (besides enhanced stimulus-seeking tendencies) for predicting that they will perform well in psi tasks.

More generally, psi researchers typically begin with the working assumption that whatever its underlying mechanisms, psi should behave like other, more familiar psychological phenomena and that psi performance should covary with experimental and subject variables in psychologically sensible ways (Schmeidler, 1988). This assumption is central to the research reported in this article, because the entire enterprise rests on the premise that time-reversed versions of well-established psychological effects will produce the same results as the standard non-psi versions and that any individual-difference correlates of psi performance, such as stimulus-seeking tendencies, will make psychological sense.

The most explicit statement of this position is embodied in the "first sight" model of psi functioning proposed by Carpenter (2004, 2005). He proposed that, far from being rare, unusual, or exotic, psi-mediated information continuously helps us to interpret experience and guide our choices in everyday life just as other preconscious information does. For example, Carpenter (2005) noted that individuals who do well at processing subliminal sensory information also do well at psi tasks (pp. 69–70). We have now seen that habituation and retroactive habituation are similarly correlated (Parker & Sjödén, 2010). The correlation between stimulus seeking and psi performance in several of our experiments also makes sense within this model. But the strongest validation of Carpenter's model is the ability of our participants to use psi information implicitly and nonconsciously to enhance their performance in a wide variety of everyday tasks.

## Psi and Physics: Metaphors, Models, and Mechanisms

The psychological level of theorizing just discussed does not, of course, address the conundrum that makes psi phenomena anomalous in the first place: their presumed incompatibility with our current conceptual model of physical reality. Those who follow contemporary developments in modern physics, however, will be aware that several features of quantum phenomena are themselves incompatible with our everyday conception of physical reality. Many psi researchers see sufficiently compelling parallels between these phenomena and characteristics of psi to warrant considering them as potential candidates for theories of psi. (For a review of theories of psi, see Broderick, 2007, and Radin, 2006.)

The development in quantum mechanics that has created the most excitement and discussion among physicists, philosophers, and psi researchers is the empirical confirmation of Bell's theorem (Cushing & McMullin, 1989; Herbert, 1987; Radin, 2006), which implies that any realist model of physical reality that is compatible with quantum mechanics must be *nonlocal*: It must allow for the possibility that particles that have once interacted can become *entangled* so that even when they are later separated by arbitrarily large distances, an observation made on one of the particles will simultaneously affect what will be observed on its entangled partners in ways that are incompatible with any physically permissible causal mechanism (such as a signal transmitted between them). The most extensive discussion of how entanglement might provide a theory for psi can be found in Radin's (2006) *Entangled Minds: Extrasensory Experiences in a Quantum Reality.* Radin argued that

> over the past century, most of the fundamental assumptions about the fabric of physical reality have been revised in the direction predicted by genuine psi. This is why I propose that psi is the human experience of the entangled universe. Quantum entanglement as presently understood in elementary atomic systems is, by itself, insufficient to explain psi. But the ontological parallels implied by entanglement and psi are so compelling that I believe they'd be foolish to ignore. (p. 235)

Bell's theorem highlights a second prominent feature of quantum theory: the role that the act of observation plays in determining what will be observed. For example, the commonsense assumption that dynamic properties of a particle (e.g., its position and momentum) have definite values before they are actually observed is falsified by the empirical confirmation of Bell's theorem. Instead, the values of such properties remain only probabilities until the act of observation "collapses the quantum wavefunction" and causes the properties to acquire definite values. Even before Bell's theorem, it was known that whether light behaves like waves or like particles depends on the conditions of observation. These features of quantum mechanics have led to "observational" theories of psi, in which it is not just the act of observation but the consciousness of the human observer that plays an active role in what will be observed (Radin, 2006, pp. 251–252).

As Radin acknowledged in the paragraph quoted above, quantum entanglement does not yet provide an explanatory model of psi. More generally, quantum theories of psi currently serve more as metaphors than models, and some psi researchers with backgrounds in physics are even more skeptical: "I don't think quantum mechanics will have anything to do with the final understanding of psi" (Edwin May, quoted in Broderick, 2007, p. 257).

**Familiarity and belief: But how can it be like that?** Among psi phenomena, precognition and retroactive influence might seem to be the most anomalous because they challenge not only our classical conceptions of space and distance, as telepathy and clairvoyance do, but also those of time and causality. Although less well known than discussions of nonlocality, alternative conceptions of time and causality also constitute an active area of discussion within physics (Barbour, 2001). An interdisciplinary conference of physicists and psi researchers sponsored by the American Association for the Advancement of Science was organized in June 2006 specifically to discuss the physics of time and retrocausation. The proceedings have been published as a book by the American Institute of Physics (Sheehan, 2006). A central starting point for such discussions is the consensus that the fundamental laws of both classical and quantum physics are time symmetric. In particular,

> They formally and equally admit time-forward and time-reversed solutions. . . . Thus, though we began simply desiring to predict the future from the present, we find that the best models do not require—in fact, do not respect—this asymmetry. . . . [Accordingly,] it seems untenable to assert that time-reverse causation (retrocausation) cannot occur, even though it temporarily runs counter to the macroscopic arrow of time. (Sheehan, 2006, p. vii)

But perhaps the most fundamental reason that precognition and retroactive influence might seem to us to be more anomalous than telepathy or clairvoyance is that we can relate the latter to familiar phenomena. That is, we have many everyday phenomena in which information travels invisibly through space. Thus, even those who are not convinced that telepathy actually exists can still readily imagine possible mechanisms for it, such as electromagnetic signal transmissions from one brain to another (which, incidentally, is not supported by the results of experiments in telepathy). Another example is provided by the apparent clairvoyant ability of migratory birds to find their way along unfamiliar terrain even at night. This ceased to be a psi-like anomaly once it was discovered that they are sensitive to the earth's magnetic field. Recently it has even been shown that the relevant sensory mechanism is located in the birds' visual systems; in a fairly literal sense, they can "see" the magnetic field (Deutschlander, Phillips, & Borland, 1999).

In contrast, we have no familiar everyday phenomena in which information travels backwards in time. This makes it difficult even to imagine possible mechanisms for precognition and retroactive influence, thereby leading to the puzzled query "But how can it be like that?"

Unfortunately, even if quantum-based theories eventually mature from metaphor to genuine models of psi, they are still unlikely to provide intuitively satisfying mechanisms for psi because quantum theory fails to provide intuitively satisfying mechanisms for physical reality itself. Physicists have learned to live with that conundrum, but most nonphysicists are simply unaware of it; they presume that they do not understand quantum physics only because they lack the necessary technical and mathematical expertise. They need to be reassured. Richard Feynman (1994), one of the most distinguished physicists of the twentieth century and winner of the Nobel Prize for his work on quantum electrodynamics, put it this way:

> The difficulty really is psychological and exists in the perpetual torment that results from your saying to yourself, "But how can it be like that?" which is a reflection of uncontrolled but utterly vain desire to see it in terms of something familiar. . . . Do not keep saying to yourself . . . "But how can it be like that?" because you will get . . . into a blind alley from which nobody has yet escaped. *Nobody* knows how it can be like that. (p. 123, emphasis added)

## On Believing Impossible Things

Near the end of her encounter with the White Queen, Alice protests that "one can't believe impossible things," a sentiment with which the 34% of academic psychologists who consider psi to be impossible would surely agree. The White Queen famously retorted, "I daresay you haven't had much practice. When I was your age, I always did it for half-an-hour a day. Why, sometimes I've believed as many as six impossible things before breakfast" (Carroll, 2006, p. 166).

Unlike the White Queen, I do not advocate believing impossible things. But perhaps this article will prompt the other 66% of academic psychologists to raise their posterior probabilities of believing at least one anomalous thing before breakfast.

## References


Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063–1087.

Barbour, J. (2001). *The end of time: The next revolution in physics.* New York, NY: Oxford University Press.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126,* 925–945.

Bem, D. J. (2003, August). *Precognitive habituation: Replicable evidence for a process of anomalous cognition.* Paper presented at the meeting of the Parapsychological Association, Vancouver, British Columbia, Canada.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115,* 4–18.

Bierman, D. J., & Scholte, H. S. (2002, August). *Anomalous anticipatory brain activation preceding exposure of emotional and neutral pictures.* Paper presented at the meeting of the Parapsychological Association, Paris, France.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106,* 265–289.

Bornstein, R. F., Kale, A. R., & Cornell, K. R. (1990). Boredom as a limiting condition on the mere exposure effect. *Journal of Personality and Social Psychology, 58,* 791–800.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49,* 229–240.

Broderick, D. (2007). *Outside the gates of science: Why it's time for the paranormal to come in from the cold.* New York, NY: Thunder's Mouth Press.

Broughton, R. S. (1991). *Parapsychology: The controversial science.* New York, NY: Ballantine Books.

Canli, T., Desmond, J. E., Zhao, Z., & Gabrieli, J. D. E. (2002). Sex differences in the neural basis of emotional memories. *Proceedings of the National Academy of Sciences, USA, 99,* 10789–10794.

Cardeña, E., Marcusson-Clavertz, D., & Wasmuth, J. (2009). Hypnotizability and dissociation as predictors of performance in a precognition task: A pilot study. *Journal of Parapsychology, 73,* 137–156.

Carpenter, J. C. (2004). First sight: Part one, a model of psi and the mind. *Journal of Parapsychology, 68,* 217–254.

Carpenter, J. C. (2005). First sight: Part two, elaboration of a model of psi and the mind. *Journal of Parapsychology, 69,* 63–112.

Carroll, L. (2006). *Alice's adventures in Wonderland and through the looking glass.* New York, NY: Bantam Dell. (Original work published 1871)

Cofer, C. N. (1965). On some factors in the organizational characteristics of free recall. *American Psychologist, 20,* 261–272.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costa, P. T. J., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI–R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Cushing, J. T., & McMullin, E. (Eds.). (1989). *Philosophical consequences of quantum theory: Reflections on Bell's theorem.* Notre Dame, IN: University of Notre Dame Press.

de Boer, R., & Bierman, D. J. (2006, August). *The roots of paranormal belief: Divergent associations or real paranormal experiences.* Paper presented at the meeting of the Parapsychological Association, Stockholm, Sweden.

Deutschlander, M. E., Phillips, J. B., & Borland, S. C. (1999). The case for light-dependent magnetic orientation in animals. *Journal of Experimental Biology, 202,* 891–908.

Dijksterhuis, A., & Smith, P. K. (2002). Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. *Emotion, 2,* 203–214.

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

Farley, F., & Farley, S. V. (1967). Extroversion and stimulus-seeking motivation. *Journal of Consulting and Clinical Psychology, 31,* 215–216.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15,* 115–141.

Feynman, R. (1994). *The character of physical law.* New York, NY: Modern Library.

Herbert, N. (1987). *Quantum reality: Beyond the new physics.* Garden City, NY: Anchor Books.

Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology, 57,* 97–114.

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

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

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Erlbaum.

Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology, 96,* 265–287.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980, February 1). Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557–558.

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

L'Ecuyer, P. (2001). Software for uniform random number generation: Distinguishing the good and the bad. In B. A. Peters, J. S. Smith, D. J. Medeiros, & M. W. Rohrer (Eds.), *Proceedings of the 2001 Winter Simulation Conference* (pp. 95–105). Piscataway, NJ: IEEE Press.

Marsaglia, G. (1995). The Marsaglia random number CDROM including the diehard battery of tests of randomness [Computer software]. Retrieved from http://stat.fsu.edu/pub/diehard/

Marsaglia, G. (1997). A random number generator for c [Computer software]. Retrieved September 29, 1997, from http://mathforum.org/kb/message.jspa?messageID=1519408&tstart=0

Parker, A., & Sjödén, B. (2010). Do some of us habituate to future emotional events? *Journal of Parapsychology, 74,* 99–115.

Radin, D. I. (1997). Unconscious perception of future emotions: An experiment in presentiment. *Journal of Scientific Exploration, 11,* 163–180.

Radin, D. I. (2006). *Entangled minds: Extrasensory experiences in a quantum reality.* New York, NY: Paraview Pocket Books.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York, NY: Appleton-Century-Crofts.

Rosenthal, R. (1990). Replication in behavioral research. *Journal of Social Behavior and Personality, 5,* 1–30.

Savva, L., Child, R., & Smith, M. D. (2004, August). *The precognitive habituation effect: An adaptation using spider stimuli.* Paper presented at the meeting of the Parapsychological Association, Vienna, Austria.

Schlitz, M., Wiseman, R., Radin, D., & Watt, C. (2005, August). *Of two minds: Skeptic-proponent collaboration within parapsychology.* Paper presented at the meeting of the Parapsychological Association, Petaluma, CA.

Schmeidler, G. R. (1988). *Parapsychology and psychology: Matches and mismatches.* Jefferson, NC: McFarland.

Sheehan, D. P. (Ed.). (2006). *Frontiers of time: Retrocausation—Experiment and theory.* Melville, NY: American Institute of Physics.

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Spottiswoode, S. J. P., & May, E. C. (2003). Skin conductance prestimulus response: Analyses, artifacts and a pilot study. *Journal of Scientific Exploration, 17,* 617–641.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2,* 105–110.

Wagner, M. W., & Monnet, M. (1979). Attitudes of college professors toward extra-sensory perception. *Zetetic Scholar, 5,* 7–17.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). London, England: Elsevier.

Wiseman, R., & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology, 61,* 197–207.

Wiseman, R., & Schlitz, M. (1999, August). *Replication of experimenter effect and the remote detection of staring.* Paper presented at the meeting of the Parapsychological Association, Palo Alto, CA.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9,* 1–29.

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

Received May 12, 2010
Revision received August 23, 2010
Accepted September 3, 2010 ■