Remembering the future: Facilitating the recall of future events

Prof. Chris Roe

Introduction

I'd like to thank the Bial Foundation for this opportunity to introduce you to a line of research in parapsychology that seems to challenge some of the assumptions we have about our relationship with time. The field of parapsychology is concerned with 'impossible things' – reported events or abilities that conflict with what the philosopher C.D. Broad (1949, p. 291) called the 'Basic Limiting Principles' of science, tenets that have been "so overwhelmingly supported by all the empirical facts … that it hardly enters our heads to question them". Broad's four principles can be summarised as follows:

- Effects cannot come before causes
- A person's mind cannot produce any direct change in the material world except those caused via the brain / sensorimotor system
- Any mental event is an event in the brain of a living body, and cannot occur in the absence of a functioning brain
- All knowledge of the world comes to us through our conventional senses or by inference from known facts

However, there is widespread belief in and reported personal experience of phenomena that prima facie are exceptions to these principles, which seems to be independent of culture, creed or historical period (e.g., Castro, Burrows & Wooffitt, 2014; Dagnall, Drinkwater, Parker & Clough, 2016). Examples of such phenomena include:

- Premonitions such as dreams that refer to (or are 'caused by') a future event
- Psychokinetic events such as the movement or distortion of objects, or the production of wellbeing changes in another organism as a result of mental intention alone
- Out of body experiences, where the centre of experience seems to be located away from the body; or near-death experiences, in which mental events seem to occur when the brain is apparently incapable of sustaining conscious activity
- Telepathic and clairvoyant experiences, in which people seem to be able to acquire information from the mind of another person or directly from the environment without the mediation of the known sensory systems

Parapsychology represents our best attempt to account for these phenomena, either in terms of existing constructs (such as misperception, errors of recall, and deception) or by invoking new constructs that can accommodate them. Given the theme of this symposium, I shall focus on the first of these, apparent violations of the cause-effect temporal relationship. With only limited space, I shall restrict myself to just one line of research that has been loosely (and rather inaccurately) labelled 'feeling the future'.

To begin with, I would like to say a few words about the originator of this line of work, Daryl Bem, who was initially scheduled to give this talk but unfortunately is unable to join us through ill health. Before he turned to parapsychology, Bem had already established an impressive reputation in more traditional areas of Psychology. Two psychological theories are named after him, accounting respectively for attributions people make to themselves to explain their own interpersonal behaviour, and for observed differences in triggers of sexual attraction. He was sufficiently well-regarded to be invited to co-author a number of editions of the standard undergraduate Psychology textbook, known colloquially by generations of students as 'Atkinson and Hilgard'. Bem was also a longstanding member of the Association of Psychic Entertainers, and it was his expertise in sleight of hand and the art of deception that led to him being exposed to parapsychological research. In 1983 Charles Honorton invited him to review the protocols he had developed to test for ESP using the ganzfeld method, to see if the security precautions could be overcome by an expert magician. Bem was sufficiently impressed that he agreed to co-author a paper with Honorton if the protocol delivered above-chance results. The findings were highly significant, and so Bem co-authored a summary report that was published in Psychological Bulletin (Bem & Honorton, 1994).

The response to this article was mixed, and suggested that some colleagues in psychology had a problem in accepting the testimony of Bem and Honorton regarding the empirical evidence they were producing, perhaps in part because the methods used to gather data were relatively unfamiliar to them. To address this, Bem set out to develop a 'Holy Grail' — "a straightforward, transparent laboratory demonstration of psi that could be replicated by any competent experimenter" (Bem, 2003, p. 6). This would involve a 'standard' psychology protocol that would be widely recognised as robust and valid, and thus in theory would be immune to methodological criticism since this would apply equally to widely accepted. It was also intended to encourage colleagues to test the claim directly for themselves rather than rely on the testimony of others; if the effects are real, then they should be reproducible in the same way as any other psychological effect. Finally, the standard protocol must involve a sequence of elements so that it could be adapted to produce a precognition design.

Bem's time-reversed protocols

Bem's first experiment of this type focused on the 'mere exposure' effect, which refers to the tendency for people to develop a preference for stimuli to which they have been exposed previously. In the mere exposure protocol, the participant would be repeatedly exposed to a stimulus and, when later presented with two stimuli and asked to make a judgement as to which is more likeable, they would tend to select the previously exposed stimulus over the novel one. The effect occurs even where initial exposure is degraded or below the level of conscious awareness (for example, with very low illumination levels, or very short exposure times); indeed, effects may be even stronger where the participant has no conscious awareness of which stimulus they have been exposed to. The effect has been described in more than 200 research articles and occurs in animals as well as humans (Monahan, Murphy & Zajonc, 2000).

Bem converted this into a precognition task by having participants make their judgement as to which of two images they preferred *before* they were repeatedly exposed to one of the two images (see Figure 1). Since the target image is selected randomly by the computer shortly *after* the participant has registered their preference, there is no obvious conventional mechanism to account for any effect. Nevertheless, Bem reported that data from more than 400 trials conducted by a number of researchers had yielded

strong support for the increased liking for negative stimuli (52.6% selection rate rather than chance expectation of 50%) that they would be exposed to in the future.

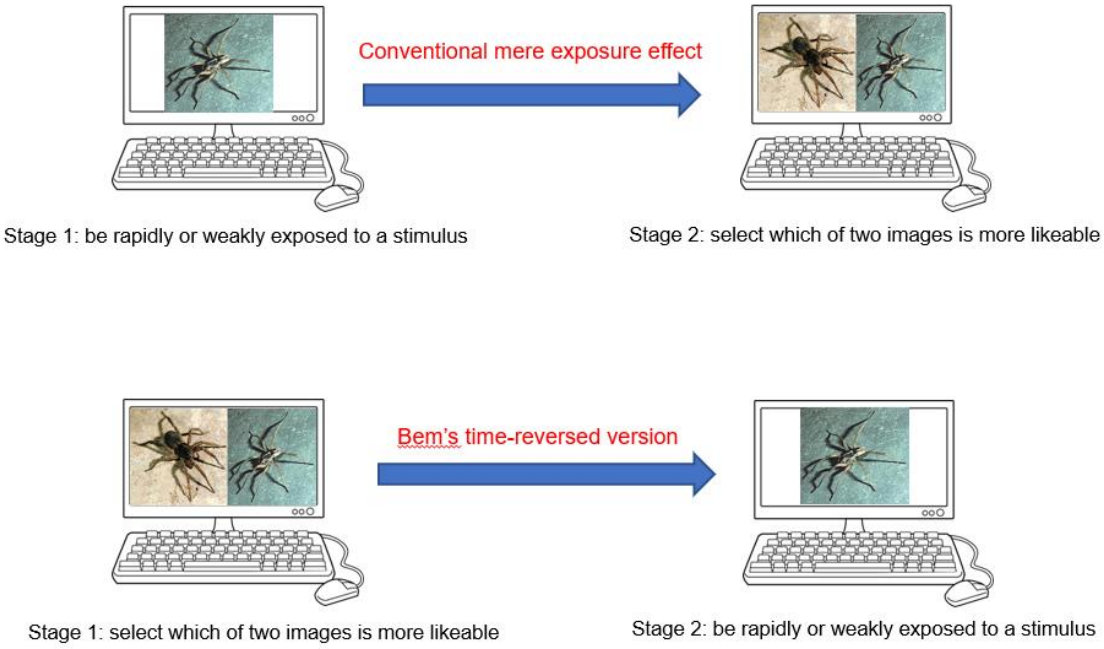

Figure 1: Conversion of the mere exposure protocol into a precognition task

Emboldened by this initial success, Bem sought other protocols that included elements that could be time reversed so as to produce a precognition task. For example, in a standard priming task the participant is presented supra-liminally with a target image and is asked to respond as quickly as possible (but without errors) to indicate whether the image is positive or negative, for example by pressing respectively a left key or a right key. Immediately before the image is presented, a positive or negative word (known as the 'prime') is presented, usually so briefly that the participant reports only seeing a flash of light rather than recognizing what word has been presented. Nevertheless, the meaning of the prime affects the participant's reaction times in responding to the image: where the prime is congruent with the overt image (for instance, the word 'beautiful' followed by an image of some flowers) the participant's reactions are typically quicker than they would be without a prime; where the prime is incongruent (for instance the word 'disgusting' followed by an image of some flowers), the correct response is slowed down relative to reaction times without a prime. Such 'semantic' priming effects are well established and considered to be robust (cf. Van den Bussche, Van den Noortgate & Reynvoet, 2009). Bem included a 'classic' priming task in his experiment, but also a condition in which the elements were reversed so that participants were presented with the image first, and *only after they had reacted to it* were they presented with a subliminal prime — by which point, of course, it would be too late for the prime to affect reaction times by any conventional means. Nevertheless, Bem reports that data from 100 participants showed that they were on average 15 milliseconds faster on congruent trials than on incongruent trials with the time-reversed (precognitive) version of the task.

A third experiment looks at the effects of practice on word recall. A staple of psychology undergraduate research methods classes, the basic effect is that participants recall more items on a list of presented words if they have had an opportunity to 'practise' them, particularly if they process them more deeply, such as by finding ways in which the words might be linked semantically, a technique known as clustering (Lockhart & Craik, 1990). This can be demonstrated by only allowing participants to practise some of the presented words and then showing that they recall more of these than the words that are presented but not practised. In Bem's time-reversed version, participants are given a chance to practise with 24 of 48 presented words, but only after they have completed the recall task. This seems akin to sitting an exam and then revising for it afterwards. Nevertheless, Bem again reports evidence of a precognition effect, with his 100 participants recalling more of the to-be-practised words than the control words.

These experiments were reported in a summary paper that was published in the *Journal of Personality and Social Psychology*, which encourages papers that report on experimental series rather than individual studies. Bem (2011) described nine formal experiments that included exploratory and confirmatory tests of the three protocols I have described, as well as some others, including a detection of erotic stimuli task that is not really a time-reversed standard protocol but is based on traditional forced choice ESP testing methods.

Reaction

Despite Bem's clear rationale in choosing experimental designs that would be more appealing to mainstream psychologists, the scientific community's reaction to the article's publication was mainly negative. A *New York Times* article (Carey, 2011) noted that "the decision may delight believers in so-called paranormal events, but it is already mortifying scientists", and quotes Ray Hyman, Emeritus Professor of psychology at the University of Oregon, "It's craziness, pure craziness. I can't believe a major journal is allowing this work in". Jarrett (2014) included Bem's study among the "10 most controversial psychology studies ever published", alongside notorious research such as Zimbardo's Stanford Prison Experiment and Milgram's "Shock Experiments". Engber (2017) described the research as "both methodologically sound and logically insane" and quotes University of Amsterdam professor Eric-Jan Wagenmakers' experience of reading Bem's ESP paper, "I had to put it away several times … Reading it made me physically unwell." In the same article, Uli Schimmack, a psychologist at the University of Toronto asserted "I don't have to believe any of these results because they're clearly fudged."

In attempting to explain this vociferous rejection of Bem's findings, Lacsap (2010) observed, "after speaking to quite a few of my colleagues about this [paper], I realize that the willingness to take these results seriously – as opposed to dismissing them out of hand – is a function … of the PRIOR probability that such effects exist… *People were bugged by the result, not the methodology*. As a matter of fact, the experimental approach (with several substudies) would have passed muster in most fields, including psychology, without a second thought if the results had been more in line with expectations. No one would have batted an eye, no one would have attempted a replication. This should give those with a concern for the state of the field pause for thought. How many results that are wrong do we believe because we expect them to be true?"

Wagenmakers Wetzels, Borsboom, and van der Maas (2011) asserted that the statistical approach adopted by Bem (and common to most psychological research) overstates the evidence against the null hypothesis, particularly where sample sizes are relatively large. They prefer a Bayesian analysis which

gives an estimate of the prior probability of a given effect and calculates how that probability shifts as a result of the observed data. Of 10 critical tests they conducted, three yielded "substantial" evidence in favour of the null hypothesis, six provided evidence in favour of an effect that was only "anecdotal", and only one (Facilitation of Recall II) gave "substantial" evidence for an effect, leading them to conclude that "Bayesian reanalysis of Bem's experiments … demonstrated that the statistical evidence was, if anything, slightly in favor of the null hypothesis" (p. 431).

Bem, Utts and Johnson (2011) responded, arguing that Wagenmakers et al. incorrectly selected an unrealistic prior distribution for their analysis and that a Bayesian analysis using a more reasonable distribution yields strong evidence in favour of the psi hypothesis. The arguments are technical, but essentially psi studies tend to give an average effect size in the range .15-.25, which is broadly comparable to effect sizes for psychology as a whole, whereas Wagenmakers et al. assumed that if the null hypothesis were false (i.e. there was a real effect size) there was more than a 50% likelihood that the effect size would be greater than 0.8. When a more realistic "knowledge-based" prior is used, five of the nine experiments gave either "strong" or "substantial" evidence in favour of an effect, and the combined Bayes factor greatly exceeds Wagenmakers et al.'s criterion for "extreme" evidence in favour of an effect.

First wave replications

Many of the concerns raised about Bem's experiments can be resolved by independent replication. A high-profile failure to replicate was reported by Ritchie, Wiseman and French (2012a). They focused on retroactive facilitation of recall, with each author overseeing an independent study involving 50 participants. All trials were conducted in-person, either by the author or a research assistant, as was the case for Bem's original experiments. All three experiments are reported to be nonsignificant; in two cases this is because the mean difference in recall for practice words and control words is very small, but replication 2 gives a 1-sample t-test value of 1.57, which is a suggestive effect. The authors regard this as nonsignificant because it is in the opposite direction to prediction (participants recalled more of the control words than practice words) and so would be rejected by a 1-tailed test. However, it seems an odd decision to adopt 1-tailed tests given that they echo criticisms of Bem for using them, especially when experimenter effects linked to their scepticism of psi (versus openness to it) have been observed for other psi experiments – see Roe (2016) for a fuller consideration. The uncorrected weighted mean recall score gives $t = 3.09$, which for a sample of 50 participants would give a highly significant ($p < .005$) missing effect even if corrected for multiple analyses. Nevertheless, it is clear that none of these replication attempts confirmed Bem's original findings.

Ritchie et al. submitted a report for publication in the *Journal of Personality and Social Psychology*, but were surprised when it was rejected. They attributed this to journal editors having little appetite for publishing failures to replicate, though the journal also rejected submissions that claimed to support the Bem findings (Aldhous, 2011). While antipathy for null results may be generally true in the social sciences, and is likely to have had an impact on the published record as a whole by skewing it to the positive (see, for example, Schmidt, 2009), it is a surprising attribution to make in this case. The article seems unlikely to meet the journal's criteria that submissions will be evaluated on the basis of the statistical power of the study that is carried out, and the number and power of previous replications of the same finding. In this case, three low powered experiments have little prospect of providing an adequate refutation of the original studies. I conducted a power analysis to estimate the likelihood that a study with sample size 50

would produce an outcome that was significant at p = .05 (1-tailed) given an effect size d of .19 (as reported in Bem's experiment 8), and this produced a power estimate of .37. In other words, there is only a 37% chance that an individual study would successfully replicate the original significant outcome where the effect is real but small. A simple binomial analysis indicates that a collection of three such studies would all be nonsignificant about 25% of the time. However, if we use the much larger effect size d = .42 from Bem's experiment 9, then the power of each replication attempt increases markedly to 90% and the likelihood that none of the 3 is independently significant reduces to 0.1%. Nevertheless, the authors attracted a lot of media attention that was sympathetic to the claim that their initial publication difficulties were due to the mistreatment of failed replications, featuring for example in articles in *New Scientist* (Aldhous, 2011), *The Guardian* (French, 2012) and even in the *Stanford Encyclopedia of Philosophy*'s entry on 'Reproducibility of Scientific Results' (Fidler & Wilcox, 2018). The British Psychological Society's professional member magazine, *The Psychologist*, devoted an issue to concerns about replication that was opened by a summary of the Ritchie et al. replication failure (Ritchie, Wiseman & French, 2012b).

With respect to statistical power issues, Galak, LeBoef, Nelson and Simmons (2012) conducted a much more substantial replication attempt, comprising seven experiments and over 3,000 participants. This focused on Bem's facilitation of recall effect on the reasonable grounds that "the other findings reported in Bem (2011) hinge on nuanced affective responses" that can be "be sensitive to subtle variation in the intensity and character of the stimuli" (p. 934). In contrast, with the selected experiment, participants are simply shown a list of words in the knowledge that they will subsequently be asked to recall as many as they can. This series of experiments adheres broadly to Bem's approach but does incorporate changes; for example, experiments 1, 2, 6 and 7 were conducted online, experiment 2 used (unspecified) different words and different categories, and experiment 6 included a 'standard' recall task. Participation pathways to the online experiment 7 — by far the largest study — included hyperlinks from an online report that described the original Bem study. It is not clear whether participants recruited by this method might have thereby been introduced to the set of test words used in the original study and (presumably) re-used here. Sample sizes for the seven experiments are very uneven at 112, 158, 124, 109, 211, 175 and 2,469, and this variation is not adequately explained.

Only one of the seven experiments — experiment 4 — showed a significant effect suggesting precognition (using a one-tailed p value), and the combined effect across all studies was very close to zero. Interestingly, the three experiments conducted in-person gave t values of +1.28, +1.77, and -0.71 (for comparison, Bem's original recall experiments gave t values of +1.92 and +2.96), whereas the online experiments gave t-values of -1.20, 0.00, -.33 and -.23. Considering just the in-person experiments gives a positive but non-significant effect size of 0.07 (Z = 0.940, p = 0.347).[1] It seems as if adopting an online protocol is not a valid variation. Online research clearly has a number of advantages, particularly with respect to generating large samples of participants and enabling people to participate at times that are convenient for them. However, there are marked disadvantages that result from participation not being monitored at any level by an experimenter: there are no checks that participants are attending to the task to the exclusion of all distractions; it is not possible to verify that participants are not cheating by writing down the words as they appear; and there is no facility to check whether participants are selectively withdrawing from the study (for example, if they discover that the words they have recalled are not among the words they subsequently have a chance to practise). To their credit, Galak et al. attempted to gauge participant

[1] With thanks to Patrizio Tressoldi for calculating these statistics.

attentiveness, but the approaches they incorporated (to ask people if they had been attentive, and to measure how long it took to complete the task) seem naïve and crude respectively. Until more effective methods have been built into their designs, data collected online is likely to remain of dubious validity.

Galak et al. (2012) additionally presented a meta-analysis of all replication attempts to date, including their own suite of experiments and the replication failures by Ritchie et al. (2012a). All studies in the database involved the facilitation of recall effect, and all were in-person tests apart from the four experiments by Galak et al. described above. The overall average effect size of .04 is considerably smaller than Bem's (2011) average effect size (.29) and is not statistically different from zero. This effect size is weighted by sample size and may have been disproportionately affected by Galak et al.'s experiment 7, which had 2,469 participants (over 60% of the total). This study was online and so suffers from the potential problems I have outlined; interestingly, a separate analysis by Galak et al. (2012) that excluded all online experiments gives a significant effect size of .09.

Bem, Tressoldi, Rabeyron and Duggan (2016) published a more comprehensive meta-analysis that encompassed all the 'feeling the future' protocols to date. They retrieved 69 attempted replications as well as 11 other experiments that "tested for the anomalous anticipation of future events in alternative way". If Bem's original studies are included, the total sample comprises 90 experiments from 33 different laboratories located in 14 different countries, and involved 12,406 participants. The replications should resolve some of the controversy surrounding Bem's original work, since they were designed from the outset as confirmatory studies that were constrained to test for the specific effects described by Bem (2011) – 31 studies are described as "exact replications" and 38 as "modified replications". The overall effect size (Hedges' g) is 0.09, which is significant ($p = 1.2 \times 10^{-10}$) and is interpreted by the authors as "decisive evidence for the experimental hypothesis" (p. 7). When Bem's original experiments are removed from the analysis, the result remains highly significant. There were differences in outcome across experiment-types, with "fast-thinking protocols" which require quick judgements that do not allow time for reflection (such as the priming task) producing larger effects than the "slow-thinking protocols" (such as memorising and recalling words). It is interesting to note that the flurry of failures to replicate Bem's findings had all focused on this latter task type. Concerns about selective reporting are tested by comparing outcomes from peer reviewed publications with 'unpublished' studies (including conference proceedings); these did not differ in outcome, suggesting there was no overt publishing bias. It is possible to calculate the number of unpublished studies that average a null result which would be needed to cancel out the observed effect; in this case there would need to be more than 1,000 unpublished experiments, which is extremely unrealistic.

As a postscript I should note three unsuccessful priming experiments that have been reported since this second meta-analysis was published (Schlitz et al., 2021; Schlitz & Delorme, 2021). These experiments involved healthy sample sizes (N = 495, 564, and 246) and were conducted in-person rather than online, so it is unclear why there was no effect in the primary measure. Schlitz et al. (2021) recruited 32 experimenters who in turn recruited their own participants. As well as finding no overall priming effect, there was no effect of experimenter prior belief or expectancy. However, an exploratory analysis did find a priming effect in the English language version only, which seems to be related to sampling rather than language per se (the English language sample scored higher on 4 of 5 pre-specified predictor variables: practice of a mental discipline, belief in ESP, personal experience, and being easily bored). The second study involved participants reading a pro-psi or anti-psi statement before completing the trials. This study

also failed to replicate the priming effect but did find significantly better performance from participants who received the pro-statement.

Conclusions

In summary, Bem's original proposal to adapt well-established psychology protocols so that they become a test for precognition is laudable. While it did not protect the work from methodological criticism, it has encouraged a range of researchers who would not normally get involved in parapsychological research to conduct replication attempts. Bem's stimulus paper has been widely criticised. Some of these criticisms are without merit, but others have legitimately drawn attention to inconsistencies and ambiguities in how the original studies were conducted and organised. Particular concerns around differentiating between exploratory and confirmatory studies are effectively resolved by the occurrence of independent replication attempts. Popular attention seems to have given a surprising amount of weight to three small but high-profile replication attempts. A more sophisticated understanding of the relationship between effect size and study power could have led to a more realistic understanding of the likelihood of achieving statistical significance where one is testing for a small but robust effect. An initial meta-analysis suggested that Bem's claimed effects could not be replicated, but these seem to have been compromised by the inclusion of online experiments with extremely large sample sizes that dominate that analysis. A more recent meta-analysis claims that effects can be replicated statistically and provides useful indicators for the next wave of replication attempts, particularly to map and explain the apparent advantage of fast-thinking protocols. While data from these second wave replications are encouraging, the effects remain small and precarious. Until we are able to identify necessary and sufficient conditions to produce a more robust and higher-yield effect it remains premature to speculate on the implications of these time-reversed effects for our understanding of time

References

Aldhous, P. (2011). Journal rejects studies contradicting precognition. *New Scientist*, 5 May 2011. Available at https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/

Broad, C.D. (1949). The relevance of psychical research to philosophy. *Philosophy*, *24*, 291-309.

Bem, D.J. (2003). Precognitive habituation: Replicable evidence for a process of anomalous cognition. *Proceedings of Presented Papers: The Parapsychological Association 46th Annual Convention*, 6-20.

Bem, D.J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407-25.

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

Bem, D.J., Tressoldi, P.E., Rabeyron, T., & Duggan, M. (2016). Feeling the Future: A Meta-Analysis of 90 Experiments on the Anomalous Anticipation of Random Future Events. F1000Research, doi: 10.12688/f1000research.7177.2

Bem, D.J., Utts, J., & Johnson, W.O. (2011). Must psychologists change the way they analyze their data? Journal of Personality and Social Psychology 101, 716-19.

Carey, B. (2011). Journal's Paper on ESP Expected to Prompt Outrage. Benedict Carey. *The New York Times*, Jan. 5. Available at https://www.nytimes.com/2011/01/06/science/06esp.html

Castro, M. A., Burrows, R. and Wooffitt, R. (2014). The paranormal is (still) normal: The sociological implications of a survey of paranormal experiences in Great Britain. *Sociological Research Online*, *19*(3), http://www.socresonline.org.uk/19/3/16.html

Dagnall, N., Drinkwater, K., Parker, A., & Clough, P. (2016). Paranormal experience, belief in the paranormal and anomalous beliefs. *Paranthropology*, *7*(1), 4-14.

Engber, D. (2017). Daryl Bem proved ESP is real. Which means science is broken. *Slate Magazine*, May 17.

Fidler, F., & Wilcox, J. (2018). Reproducibility of Scientific Results. Stanford Encyclopedia of Philosophy. Available at https://plato.stanford.edu/entries/scientific-reproducibility/

French, C. (2012). Precognition studies and the curse of the failed replications. The Guardian, 15 March. Available at https://www.theguardian.com/science/2012/mar/15/precognition-studies-curse-failed-replications.

Galak, J., LeBoef, R., Nelson, L., & Simmons, J. (2012). Correcting the past: Failures to replicate psi. Journal of Personality and Social Psychology 103, 933-48.

Jarrett, C. (2014). The 10 Most Controversial Psychology Studies Ever Published. British Psychological Society research Digest, https://digest.bps.org.uk/2014/09/19/the-10-most-controversial-psychology-studies-ever-published/

Lascap, [no initial] (2010). Is reality retrocausal? *Pascal's Pensées*. Available at https://blog.pascallisch.net/is-reality-retrocausal/

Lockhart, R.S., & Craik, F.I.M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, *44*(1), 87-112.

Monahan, J.L., Murphy, S.T., Zajonc, R.B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science*, *11*, 462-66.

Ritchie, S.J., Wiseman, R., & French, C.C. (2012a). Failing the future: Three unsuccessful attempts to replicate Bem's 'retroactive facilitation of recall' effect. *PLOS One*. https://doi.org/10.1371/journal.pone.0033423

Ritchie, S.J., Wiseman, R., & French, C.C. (2012b). Replication, replication, replication. *The Psychologist* 25(5), May, 346-48.

Roe, C.A. (2016). Experimenter as subject: What can we learn from the experimenter effect? *Mindfield*, *8*(3), 89-97.

Schlitz, M., Bem, D., Marcusson-Clavertz, D., et al. (2021). Two replication studies of a time-reversed (psi) priming task and the role of expectancy in reaction times. *Journal of the Society of Scientific Exploration*, *35*(1), 69-94. https://doi.org/10.31275/20211903

Schlitz, M., & Delorme, A. (2022). Examining implicit beliefs in a replication attempt of a time-reversed priming task. *F1000Research*. https://doi.org/10.12688/f1000research.27169.2

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90-100.

Van den Bussche, E., Van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, *135*(3), 452-77.

Wagenmakers, E-J., Wetzels, R., Borsboom, D., & van der Maas, H.L.J. (2011). Why psychologists must change the way they analyze their data: The case of psi — Comment on Bem (2011). Journal of Personality and Social Psychology 100, 426-32.