NOT FEELING THE FUTURE: A FAILED REPLICATION OF RETROACTIVE FACILITATION OF MEMORY RECALL

by ERIC ROBINSON

ABSTRACT

The present paper reports an attempted replication of a time-reversed psi effect reported by Bem (2011), in which word recall during a memory test appeared to be facilitated retroactively. In the experiment, immediately after completing a recall test of 48 words, participants were exposed to 24 of the originally tested words and instructed to 'rehearse' them. In the original experiments participants recalled a greater proportion of the 'rehearsed' words compared with control words, suggesting retroactive facilitation of memory recall. In the present study, fifty undergraduate participants completed the experiment, using the same software as reported by Bem (2011). A personality measure of 'stimulus-seeking' was also reported to predict the effect. The present study ($N = 50$) failed to find evidence of retroactive facilitation of recall or any effect of stimulus-seeking on performance. There was no observed difference in the number of practice and control words recalled ($M = 9.4$ and 9.8 respectively; $t(49) = 0.85$, $p = 0.40$), nor did stimulus-seeking predict retroactive facilitation of memory recall; $r(50) = -0.07$, $p = 0.65$. Further replication of the original effect is needed to substantiate claims of retroactive influence using this paradigm.

INTRODUCTION

The proposed time-reversed form of psi, commonly known as 'precognition', has been the basis of empirical study using a number of paradigms (Krippner, Ullman & Honorton, 1971; Radin, 2004; Steinkamp, Milton & Morris, 1998). Recently Bem (2011) reported a collection of nine studies that he suggested provides strong evidence for retroactive influence on human cognition. This claim is based upon the studies appearing to show that later exposure to a stimulus had a profound effect on earlier responses. In all of these experiments Bem effectively time-reversed the order of well-known psychological effects. For example, in one experiment Bem reversed the 'mere exposure' effect; whereby subliminal exposure to a stimulus results in a greater liking of the stimulus (Zajonc, 1968). In Bem's experiment participants selected which of two stimuli they preferred and then after their choice one of the stimuli would be randomly selected and subliminally presented on screen. Results indicated that participants were significantly more likely to choose a stimulus if it was later presented (Bem, 2011). The nine experiments are particularly impressive in both their level of experimental control and consistency of findings.

One reason Bem designed such experiments was an attempt to produce compelling evidence for psi that could be easily replicated by other researchers (Bem, 2003). Indeed, if one is to accept the premiss that human cognition can be retroactively influenced, consistent replications of such findings are needed. If the experiments reported by Bem can be easily and reliably replicated, then parapsychology may well have discovered its much-needed 'holy grail', so to speak. Yet, although some attempted replications have been presented

at conferences, few replication attempts are recorded in the peer-reviewed literature.

Savva, Roe and Smith (2006) report two studies which showed mixed results using an adaptation of the retroactive habituation experiment (Experiment 6 in Bem, 2011), with one study finding a small overall effect and the second finding no psi effect using a larger sample. More recently Parker and Sjödén (2010) failed to find evidence for retroactive habituation in their overall sample, but reported some findings of interest that were suggestive that the tendency to exhibit habituation of stimuli under normal conditions (i.e. in the experiment in normal time order) predicted psi performance when the order of the experiment was reversed. Of the nine experiments, Bem suggested that the retroactive priming and retroactive facilitation of memory recall studies would be the best candidates for replication (Bem, 2011).

The present study attempted to replicate one of these studies which appeared to show evidence for retroactive facilitation of memory recall (Experiment 8 in Bem, 2011). In this experiment participants are presented with 48 words and shortly afterwards asked to complete a recall test of the words. To examine a case for retroactive facilitation, participants are exposed to 24 randomly selected words from the originally tested words; 'the practice words' and are instructed to rehearse them (the remaining 24 words act as control words). Bem hypothesised that if memory can be retroactively influenced, participants should recall more 'practice words' than controls. In the original sample of 100 psychology undergraduates, the results were in line with this hypothesis. Furthermore, a personality measure of 'stimulus-seeking' predicted the effect, whereby only high scorers on the measure appeared to show the effect (Bem, 2011). This finding is in line with previous research that has linked extraversion and psi performance (Honorton, Ferrari & Bem, 1992). In an attempt to produce a close-to-exact replication of the originally-reported experiments, the present study used the same software as reported in Bem (2011) and the same measure of stimulus-seeking.

METHOD

Participants

Fifty psychology undergraduate students studying at the University of Liverpool (35 female, 15 male; mean age = 21.2 years, $SD = 3.4$) participated in the study in exchange for course credit, signing up for a session through poster advertisement. Prior to the study participants were not given any information concerning the aims of the study, only that it involved the investigation of ESP. Participants gave informed signed consent and the study protocol was approved by a University of Liverpool Research Ethics Committee.

Materials

The experiment software was identical to that utilised by Bem (see Bem, 2010 for further information). Experimental sessions were run on a Windows 98 desktop computer.

Stimulus-Seeking Questions

Participants completed two items to measure stimulus-seeking. Q1. "In general, I am easily bored", Q2. "I often enjoy seeing movies I've seen before".

Both questions were answered using five-point Likert scales (anchors: 'strongly disagree' and 'strongly agree'). Scoring was reversed for Q2.

Procedure

On arrival participants were seated in a computer laboratory and informed that the study investigated ESP. Before leaving the lab the experimenter asked participants to turn off their mobile phones and to follow the on-screen instructions. Participants were first instructed to provide demographic information, before completing an eight-item questionnaire that included the two stimulus-seeking questions. Participants then were informed that a number of common nouns would be presented on screen and their task was to visualise the referent of each noun (for example, if the word was 'tree', they should visualise a tree). Forty-eight nouns were presented on screen individually for three seconds each. Participants were then given a surprise free recall test of the 48 nouns and were asked to type their answers on screen, for which they were given a five-minute time limit.

The software then randomly selected 24 of the nouns to serve as 'practice words' (the remaining 24 served as 'control words'). The practice words were then presented in a randomised list. Participants were then informed that the words in the randomised list were selected from four categories (foods, animals, occupations and clothes). Participants were then presented with one of the four categories (e.g. foods) and asked to click on all items relating to that category and retype these words. This procedure was repeated for the words belonging to the remaining three categories. Thus, for the 'practice words', each participant scanned the scrambled list of words several times and located, clicked on and retyped each individual word.

Incorrect Spellings

Due to the possibility of participants misspelling a word, the software allows experimenters to change any obvious incorrect spelling mistakes participants have made in the free recall test. To ensure that performance would not be penalised for incorrect spelling, the experimenter made any changes to misspelt words immediately after each session.

ANALYSIS

The analysis reported here is the same analysis as reported by Bem (2011). In the free-recall test each word the participant recalls constitutes a trial and is scored as either a practice word or a control word. To measure retroactive influence on memory the number of control words recalled is subtracted from the number of practice words for each participant (P–C). However, as the participants vary in the total number of words they recall (and therefore in the number of trials they contribute to the experiment), the P–C difference is weighted in analysis. Thus, the following formula is used:–

$$(P - C) \times (P + C)$$

For continuity purposes with Bem's original paper, each participant's score is expressed as a percentage of the maximum possible score that could be achieved (based on participants recalling all the practice words and none of the control words; i.e. P = 24; C = 0). Thus:–

$$\text{Performance score (\%)} = [(P - C) \times (P + C)] \times 100\% / 576$$

Positive percentage scores indicate more practice words being recalled than control words and negative percentage scores indicate that more control words were recalled than practice words. If practising a set of words after the recall test does retroactively facilitate recall of these words, we would expect to see a positive percentage score that is significantly greater than zero. To examine this, a one-sample *t*-test was used.

To examine the effect of stimulus-seeking participants' scores on the two questions were averaged. Participants scoring above the mid-point of the scale were classified as high in stimulus-seeking and those scoring at or below the mid-point were classified as low in stimulus-seeking. An independent-samples *t*-test was used to examine whether the two groups differed in their percentage performance scores. Stimulus-seeking and performance score were also correlated using Pearson's *r* to examine any relationship between the two variables.

RESULTS

In the overall sample participants recalled an average of 19.2 words, $SD = 3.4$ (for practice words, $M = 9.4$, $SD = 3.0$; for control words, $M = 9.8$, $SD = 3.1$). The average performance score was –1.6% ($SD = 13.0\%$), which was not significantly different from chance expectation and in the opposite direction to the hypothesised effect [$t(49) = 0.85$, $p = 0.40$].

No significant correlation was observed between test performance and stimulus-seeking [$r(50) = -0.07$, $p = 0.65$]. Participants defined as high in stimulus-seeking (n = 20) recalled 9.1 practice words ($SD = 2.5$) and 9.7 control words ($SD = 2.9$), resulting in a score of –2.3% ($SD = 14.2\%$). Participants defined as low in stimulus-seeking ($n = 30$) recalled 9.5 practice words ($SD = 3.3$) and 9.9 control words ($SD = 3.2$), resulting in a score of –1.1% ($SD = 11.3\%$). There was no significant difference between the two groups' performance scores [$t(48) = 0.32$, $p = 0.75$].

DISCUSSION

The present study failed to find any evidence for retroactive facilitation of memory recall. Recall of subsequently 'practised' words was not any different from recall of control words that were not practised (and close to being exactly the same). In Bem (2011) participants classified as being high in stimulus-seeking appeared to be responsible for the overall effect in the study's sample. Here no evidence of stimulus-seeking predicting the hypothesised psi effect was found. Why Bem's (2011) reported effect was not found here is open to question. Although we used a smaller sample than the original, our results clearly show that statistical power is not responsible for our null findings. The data do not even hint at any form of retroactive psi effect. Both test performance in the overall sample and participants high in stimulus-seeking fail to differ from chance expectation, and close examination of the means in the data set shows that participants recalled a slightly smaller number of practice words than control words.

Although we used close to the same procedure as Bem, participants did not undergo what is described in the Bem (2011) paper as a "relaxation period",

during which participants watched a screen-saver-like display and listened to 'new age' music for three minutes before starting the experiment. Whether this relaxation period is of importance or has any effect on mood is unclear. Although mood has been suggested to influence performance on explicit psi tasks (Schmeidler, 1971), this experiment is somewhat different from an explicit psi task. Due to practical issues we did not include the relaxation period, but further research examining whether it has any effect on performance would be of interest. Aside from the relaxation period, the present study does not appear to differ from that of Bem (2011). Both used an undergraduate population, although our participants were UK-based. Participants in the present study recalled a similar number of words to participants in the original Bem (2011) experiment (19.2 vs. 18.4) and the proportion of high vs. low stimulus-seeking participants is similar across both studies. One possible factor may be cultural differences in word usage, whereby the test words used are common in the US but not the UK, which may somehow impede performance. This seems unlikely, as the majority of words used in the experiment do not appear to be uncommon in the UK, and such an effect would presumably reduce performance for both control and experimental word recall, which would leave psi performance scores unaffected.

A further point of interest for future work is whether better measures of stimulus-seeking may be required. The measure used in Bem's analysis (and here) is self-devised and consists of only two items. Examining the correlation between responses on the two questions here *post-hoc* suggests only a weak association; $r(50) = 0.32$, $p = 0.03$. If stimulus-seeking is of interest for further research, a more psychometrically robust measure may be needed.

As this is one of few published replication attempts, further studies are needed in order to examine whether the 'holy grail' of an easily and reliably replicated psi experiment has been found or whether the original reported results are a statistical artefact. Although the large number of experiments reported in Bem (2011) make the latter seem unlikely, the failure here to replicate what was described as one of the best candidates for replication suggests that further work is needed.

ACKNOWLEDGEMENT

I would like to thank Daryl Bem for providing the software for the study.

School of Psychology
University of Birmingham
Birmingham B15 2TT

REFERENCES

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,* 407–425.

Krippner, S., Ullman, M. and Honorton, C. (1971) A precognitive dream study with a single subject. *JASPR 65,* 192–203.

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

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

Savva, L., Roe, C. A. and Smith, M. D. (2006) Further testing of the precognitive habituation effect using spider stimuli. *JSPR 70,* 225–234.

Schmeidler, G. R. (1971) Mood and attitude on a pretest as predictors of retest ESP performance. *JASPR 65,* 324–335.

Steinkamp, F., Milton, J. and Morris, R. L. (1998) A meta-analysis of forced-choice experiments comparing clairvoyance and precognition. *JP 62,* 239–246.

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