

## Sensing the Future: Reversed Causality or a Non-standard Observer Effect?

Eugene Subbotsky\*

Lancaster University, UK

**Abstract:** In three experiments, Daryl Bem's experimental method on retroactive facilitation of recall was replicated, with minor modifications, with the aim of testing robustness of Bem's retroactive effect. In Experiment 1, the original double-blind procedure was replaced by the single blind procedure under which the experimenter, but not participants, knew what the experimental conditions were. In the experimental condition of this experiment Bem's results were successfully replicated, however, in the control condition, which was supposed to produce null effect, the effect was significant and a mirror image of the effect in the experimental condition. In Experiment 2, the control condition was run before the experimental condition. As in Experiment 1, the results in both conditions were symmetrical, though this time not significantly different from zero. In Experiment 3, in the experimental condition the results were reversed to those in Experiment 1. Altogether, the results marginally support robustness of Bem's reported effect. However, the results do not support the interpretation of this effect as an effect in inverted causality. Rather, the results of this study suggest that the effect reported by Bem is a particular case of a number of possible outcomes, which occur when Bem's method is applied. The results also suggest that these outcomes occur due to a direct effect of the observer's mind on the RNG functioning.

**Keywords:** Observer effect, ESP task, memory, reversed causality, dissipative system, oscillating pattern, symmetry, decline effect.

In his paper "Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect" psychologist Daryl Bem reported nine experiments on the precognition of future events that could not be accessed through any known physical processes or anticipated by any known inferential processes [1]. Precognition is a class of *extrasensory perception* (ESP) – the ability to access information through time or space by means unknown to modern science. In these experiments, some well-established psychological effects, such as priming, avoidance of negative stimuli, and habituation were "reversed in time", so that participants' responses were obtained before the occurrence of the causal stimulus events. Accordingly, Experiment 8 drew on the known effect in memory, according to which repetition and practice enhances recall (see [2]): the hypothesis was tested that this effect can work "backwards", by examining whether rehearsing a set of words makes them easier to recall even though the rehearsal takes place after the recall test is given. In this experiment, university undergraduates were first shown a set of words on a computer screen and then unexpectedly given a free recall test of those words. Immediately after the test, participants were given a set of practice exercises on a randomly selected subset of those words that comprised half of the original words. The remaining words

served as control, not-practiced words. The hypothesis was that the practice exercises would retroactively facilitate the recall of practiced words, so that more practiced than not-practiced words would be recalled. The procedure for the control condition in the experiment strictly followed the procedure of the experimental condition, with the computer randomly selecting the practice set of words, but not actually administering the practice exercises. This condition was necessary to find out whether the practice exercises in the experimental condition retroactively enhanced overall recall. The second aim of the control condition was to confirm that it is practicing with practice words that produces their enhanced recall, not just the bare fact of the computer randomly choosing a set of practice words which participants might be able to access *via* ESP<sup>1</sup>.

The results of the experiment showed that in the experimental condition, practiced words were recalled significantly more frequently than not-practiced words, suggesting that practicing after the test did, indeed, enhance recall. Interestingly, in both experimental and control conditions overall recall averaged 18.4 (38.3%) of the 48 words presented in the original list. Bem noted that this conformed to the effect

\*Address correspondence to this author at the Psychology Department, Lancaster University, Lancaster LA1 4YF, UK; Tel: 0152464210; Fax: 01524593744; E-mail [e.subbotsky@lancaster.ac.uk](mailto:e.subbotsky@lancaster.ac.uk)

<sup>1</sup>In my view, the control condition also safeguards against the possibility that the retroactive facilitation of recall is an artefact of the computer choosing the set of practice words, which, due to some malfunction of the selection procedure, might have been influenced by the words that the participant had recalled.

known in memory studies as “retrieval-induced forgetting”, according to which practicing items to be recalled increases the likelihood that these items will be recalled at a later time but also causes forgetting of other items associated to the same cues guiding retrieval that were not practiced [3]. Thus, it is possible that the enhanced recall of practiced words came at the expense of diminished recall of not-practiced words. In the control condition, as predicted, there was no significant difference between the frequencies of recalled practiced and not-practiced words. Experiment 9 reported in Bem’s paper was a replication of Experiment 8, with the exception that an additional practice exercise was introduced, in an attempt to enhance the retroactive facilitation effect. The results yielded an even stronger facilitation effect than in Experiment 8.

One specific feature of Bem’s experiments was that the retroactive facilitation of recall effect (as, in fact, other retroactive effects reported in his paper) did not have an explanation that could fit this effect to the existing frame of views about how brain neural mechanisms and psychological processes work. Indeed, it is commonly accepted that training on memorizing a certain material makes this material easier to recall, and this notion fits well with the general scientific outlook, i.e., with our knowledge about brain mechanisms that underlie memory (i.e., strengthening neural pathways in the brain with repetition), or with the general causality principle that the effect (the easier recall) follows the cause (training) and not vice versa. Clearly, training that assists recall “from the future” violates this outlook.

Yet, in his paper Bem suggested that the reported effect is a version of a regular psychological effect and can be replicated. But is the effect, which violates the fundamental principle of causality, on which the whole body of modern science rests, a regular psychological effect? If it is, then it should possess the features, which regular psychological effects possess. One of such features is replicability. Two recent attempts to replicate Bem’s retroactive facilitation effect without changes, by exactly following Bem’s methodology, have failed to replicate this effect [4, 5]. Another feature of a regular psychological effect is robustness, which implies replicability of an effect with minor changes both in the effect and in methodology. Indeed, a regular psychological effect must be robust, which means that, with minor changes in the methodology, this effect should not disappear or alter in some major way: instead, the effect should replicate with minor alterations (i.e., with a larger or smaller effect sizes). For example, it was shown that varying types of repetition of a category name and items that presented instances of that category (i.e., precious stones: Rubi, Opal) changes the size of the facilitative effect, yet all types of repetition produced a large facilitative effect on recall of the basic unit [6]. While replicability and robustness are linked to each other, these features are not identical. There are natural and psychological phenomena, which cannot be replicated exactly, yet they can be replicated with some alterations. The most known of such phenomena is family resemblance. It is impossible that a daughter or a son, in his or her facial appearance, is an identical copy of one of their

biological parents, however, within certain limits he or she can possess a clearly identifiable family resemblance. The same effect of “robustness but not identity” is even more salient when it goes about children inheriting psychological features from their parents. It is possible that a psychological effect reported by Bem belongs to this kind of phenomena. *On this ground, the question arises, whether Bem’s reported effect, albeit it is not replicable, is nevertheless robust.*

One way of answering this question is to introduce a minor change in Bem’s methodology, for instance, by replacing the double blind procedure with a single blind one. By employing the double-blind procedure, the experiments reported by Bem are designed to avoid the observer effect, which makes these experiments particularly attractive to mainstream science. Specifically, in designing Experiment 9 of his study, Bem interspersed the control sessions among the experimental sessions: the manipulation that is usually done in order to prevent the experimenter’s expectations from affecting the results, *via* making the experimenter blind to the experimental conditions. It has to be noted that procedure of Bem’s experiments was pre-programmed on the computer and the experimenter wasn’t even visible to the participant during the experiment. It is therefore hard to explain how the experimenter’s knowledge of what condition a participant is doing – experimental or control one – might affect the results in a usual way, i.e. through behavioral cues or data adjustment. On this ground, it might be argued that interspersing control trials among experimental trials is redundant, as there is no a rational link between the experimenter’s knowledge regarding the condition and the condition’s results. Coincidentally however, this “unnecessary precaution” opens the possibility for checking whether the result of the retroactive facilitation of recall experiment is robust.

The way this check can be accomplished is through separating the control condition from the experimental one, *via*, for example, running the control condition after the experimental condition. In this case, the experiment follows the “single blind” version of the experimental procedure, whereby participants are blind as to the condition they are performing but the experimenter isn’t. If the retroactive facilitation of recall effect is indeed robust, then the results of both conditions, when they are separated, or at least in the experimental condition should be approximately the same as the results reported by Bem. If, however, the reported effect is not robust, then the shift from a double blind to a single blind procedure should produce null effect in both conditions.

In summary, the aim of the present study was to explore robustness of the retroactive facilitation of recall effect, by examining if a minor change in the procedure (i.e., replacing a double blind procedure with the single blind one) would increase, diminish or eliminate this effect altogether. Whilst strictly following the procedure of both conditions of Bem’s experiment, we separated the experimental condition from the control one, by running the control condition after the experimental condition.

**EXPERIMENT 1.**<sup>2</sup>**Method****Participants**

Participants were graduates and undergraduates at Lancaster University (UK), 75 participants (45 females and 30 males; age range 18 to 28 years,  $M = 20.39$ ,  $SD = 2.81$ ) contributed experimental sessions, and 25 (20 females and 5 males; age range 18 to 27 years,  $M = 20.00$ ,  $SD = 2.71$ ) contributed control sessions.

**Materials**

A G5 Mac computer with a 24-in. color monitor controlled the experiment, and the application program was the original one used by Daryl Bem in his Experiment 9. The application program was recompiled from the source code by Adrian Ryan, who is an experienced computer programmer and an ESP researcher. The source code was carefully scrutinized to ensure that the selection of practice words was independent of the recalled words keyed during the experiment.

**Design**

The between-subjects variables were condition (experimental and control), and sex. The dependent variables were total number of words recalled, number of practice words recalled, number of not-practice (control) words recalled and weighted differential recall score (DR%) calculated from the other dependent variables. The latter was defined, as in the original study, as the number of practice words recalled minus the number of control words recalled ( $P - C$ ) multiplied by the participant's overall recall score ( $P + C$ ) and taken as a percentage of the maximum possible DR score ( $= 576$ ), which is achieved if a participant recalled all 24 practice words and none of the 24 control words. As a result,  $DR\% = [(P - C) \times (P + C)]/576$  and can vary between -100% and 100%. The DR% score enables testing of Bem's central claim, for the existence of a retroactive effect, whilst the other, more fundamental, dependent variables enable analysis of the retrieval-induced forgetting effect.

**Procedure**

Participants were individually invited into the laboratory and given the following information: "This experiment tests for ESP (extrasensory perception) 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." As in the original experiment, participants were not told in advance about the recall test.

Participants were then invited to sit in front of the computer, told that the experiment is going to be easy and enjoyable, and asked to follow the instructions on the screen. During the experiment, the experimenter was in the same room but separated from the participant by an opaque screen. After typing in their name and age, and responding to the stimulus

seeking and transliminality questionnaires<sup>3</sup>, each participant experienced a three minute relaxation procedure during which the program played a soundtrack of New Age music while displaying moving Hubble images of the galaxy. The computer then displayed 48 common nouns one at a time for three seconds each. The words were presented in the same fixed order for all participants. The words were from four categories (food, animals, occupations, and clothes) and were picked from Francis & Kucera's [7] "Frequency analysis of English usage" handbook. In each of the four categories six low frequency and six high frequency words were used. Participants were asked to visualize the referent for each word. After all words were shown, participants were unexpectedly asked to free recall as many words as they could and type them in any order. The recall time was limited to five minutes.

After 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 not-practiced control words. The first practice exercise then followed. The participants were instructed 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 image of the thing it refers to (e.g., if the word is tree, visualize a tree)." The practice words then appeared on the screen, one by one, each for three seconds, but unlike the initial presentation, this time they were grouped in categories: six food words were followed by six animal words, and so forth. The second practice exercise followed. The 24 practice words all appeared together in a randomized list on the screen, and the participant was asked to click on the six food words in the list, which turned red when clicked, then to retype those words into six empty slots on the screen, and click the "continue" button. In the next display, the list was re-scrambled and the procedure was repeated for each of the other three categories of words. The computer then displayed a message thanking the participant and, in the experimental condition, the numbers of practiced and not-practiced words appeared on the screen. In the control condition, only the total number of words recalled was displayed. The experimenter then debriefed the participant. The order of the conditions was fixed: 75 sessions in the experimental condition were followed by the 25 sessions in the control condition.

**Hypotheses and Predictions**

The first hypothesis was an open one. If the retroactive facilitation of recall effect is indeed robust, then the results of the experimental condition in this experiment should be the same as the results of the experimental condition reported by Bem. Specifically, in the experimental condition of the retroactive facilitation of recall experiment, the recall of practice words should be significantly better than the recall of not practice words. This hypothesis will be supported if the mean DR% score in this condition is positive and significantly different from zero. The second hypothesis was derived from the the concept of robustness, which implies that

<sup>2</sup> This and subsequent experiments were run by the same male experimenter, in the same room and employed the same materials.

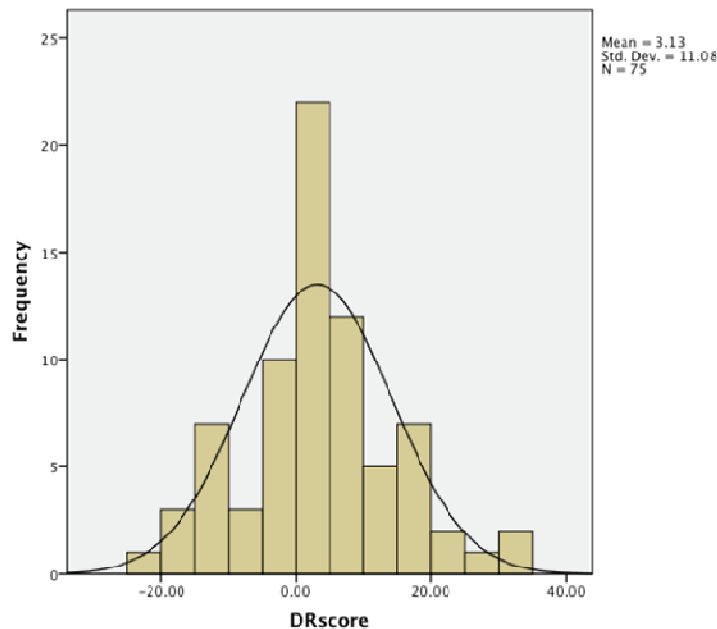
<sup>3</sup> Those were additional scales introduced by Bem in order to test some individual differences variables. As these scales had no effect on results of this experiment, they were excluded from the analysis.

small changes in the experimental methodology bring about some changes in the results, while preserving the general pattern of the main effect at the same time. Accordingly, it was expected that the effect size in the experimental condition of this experiment would differ from that reported by Bem. Finally, regarding the control condition, the expectation was again an open one. If Bem's explanation of the retroactive facilitation effect is correct, and this effect results from the practice with practice words, then in the control condition of this experiment the null effect is expected. If,

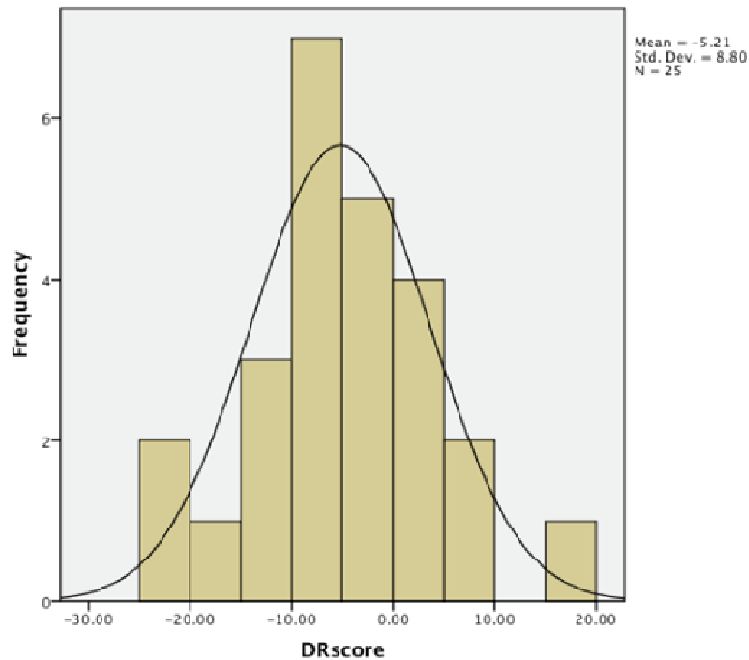
however, some factors other than practice with the practice words may have caused the better recall of practice words in the experimental condition, then in this experiment the mean DR% score of the control condition could be different from zero.

**RESULTS<sup>4</sup>**

The retroactive facilitation of recall effect<sup>5</sup>. The mean DR% score in the experimental condition was 3.13%, SD =



**Fig. (1).** Histogram of the DR% score in the experimental Condition.



**Fig. (2).** Histogram of the DR% score in the control condition.

<sup>4</sup> All significance levels reported in this article are based on two-tailed tests, with *d* and  $\eta^2$  being used as the indexes of effect size.

<sup>5</sup> As there were no significant main or interaction effects for participant sex, this variable is excluded from the analysis.

11.08. The distribution of the DR% scores was normal, Shapiro-Wilk  $W = .98$ ,  $p = .43$  (see Fig. 1). The deviation of the mean DR% score from zero was significant, one-sample  $t(74) = 2.44$ ,  $p = .017$ ,  $d = .28$ . In the control condition, the mean DR% score was  $-5.21\%$ ,  $SD = 8.80$ . The distribution of the DR% scores was also normal, Shapiro-Wilk  $W = 0.98$ ,  $p = .80$  (see Fig. 2). Again, the deviation of the mean DR% score from zero was significant, one-sample  $t(24) = -2.96$ ,  $p = .0068$ ,  $d = -.59$ .

The total numbers of words recalled. There was no significant difference between mean total number of words recalled in the experimental and control conditions,  $t(98) = -1.07$ ,  $p = .29$ .

Analysis of number of practice and not-practice words recalled. Mean numbers (SDs) of practice and not-practice words recalled as a function of condition are shown in Table 1 and illustrated in Fig. (3). A two-way ANOVA for condition (experimental and control) and type of words (practice and not-practice), with the repeated measures on the last factor, was run with number of words recalled being the dependent variable. The model showed no significant effect of type of words,  $F(1,98) = 0.92$ ,  $p = .34$ ,  $\eta^2 = .01$ . The effect of condition was not significant either,  $F(1,98) =$

$0.86$ ,  $p = .36$ ,  $\eta^2 = .009$ . There was, however, a significant interaction between condition and type of words,  $F(1,98) = 11.32$ ,  $p = .001$ ,  $\eta^2 = .10$ .

Tests of simple effects indicated that in the experimental condition the mean number of practiced words recalled was significantly higher than the mean number of not-practiced words recalled, paired samples  $t(74) = 2.33$ ,  $p = .023$ ,  $d = .29$ . Conversely, in the control condition, the mean number of practiced words recalled was significantly lower than the mean number of not-practiced words recalled, paired samples  $t(24) = -3.14$ ,  $p = .0045$ ,  $d = -.68$ .

Correlation between DR% scores and time of testing. DR% scores in the experimental condition were negatively correlated with time of testing,  $r = -.21$ ,  $p = .073$ . The highest DR% scores were achieved in the first three weeks of testing (55 participants, mean DR% score =  $4.61$ ,  $SD = 11.35$ ), after which time the DR% scores dropped sharply (20 participants, mean DR% score =  $-0.96$ ,  $SD = 9.37$ ); the difference was significant,  $t(73) = 2.14$ ,  $p = .038$ ,  $d = .53$  (see Fig. 4).

**Discussion**

The results supported the hypothesis that the effect reported by Bem is robust.

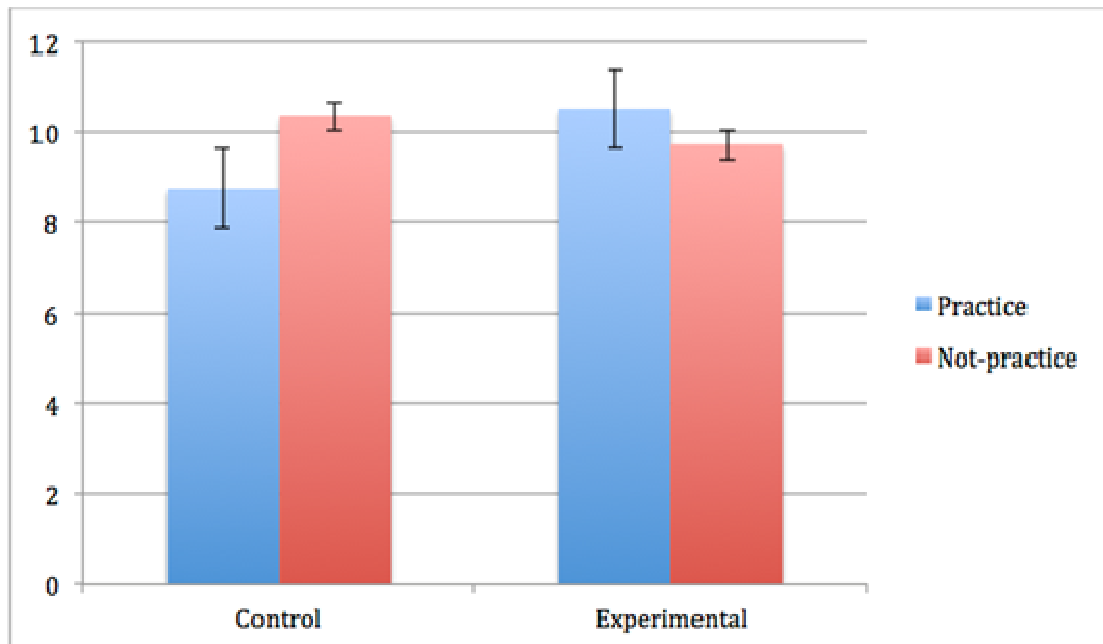


Fig. (3). Numbers of practice and not-practice words in Experiment 1.

Table 1. Words Recalled as a Function of Condition (Experimental and Control) and Type of Words (Practiced and Not-Practiced) in Experiment 1

	Control			Experimental		
	M	SD	%	M	SD	%
Practiced words	8.76	2.40	36.50	10.52	2.94	43.83
Not-practiced words	10.36	2.27	43.17	9.72	2.61	40.50
Total	19.12		39.83	20.24		42.17

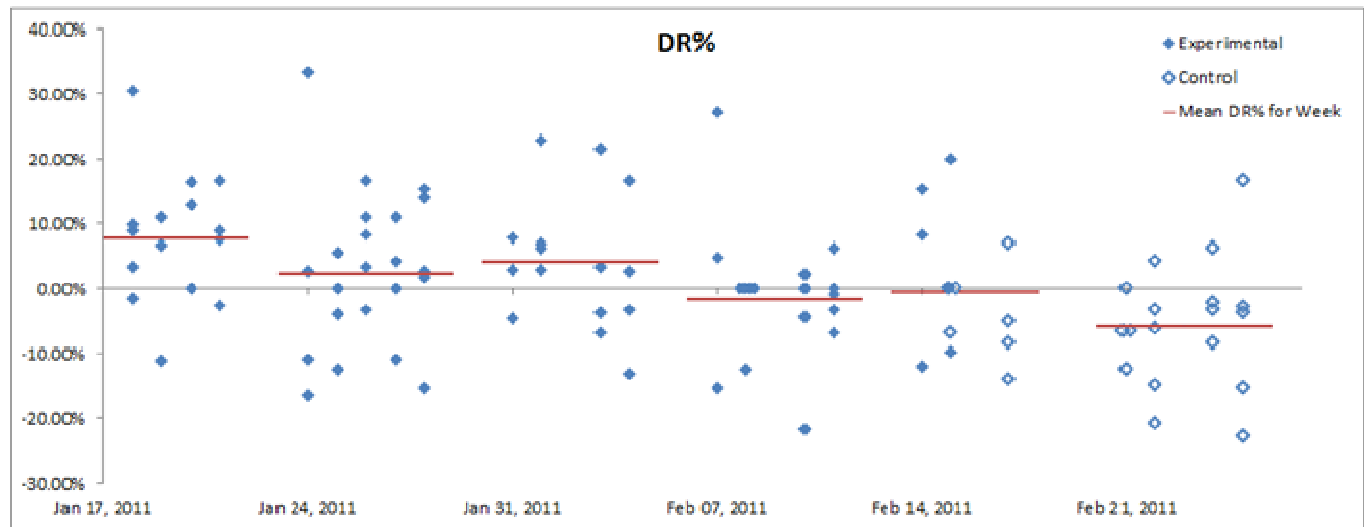


Fig. (4). DR% scores by time of testing in Experiment 1

Specifically, the experimental condition of this study replicated the results of the original study: The practice words were recalled significantly more frequently than the not-practice words. The fact that participants practiced with the words after, and not before the recall test, might suggest that the facilitating effect of practice worked “backwards in time”. The source code recompilation and audit of the application source code rules out the possibility that the observed results were due to programming error or to a bias “built-in” to the application. The mean DR% score of 3.13%,  $SD = 11.08$  was only slightly smaller than the  $M = 4.21\%$  reported by Bem (2011), and reached about the same level of significance,  $p = .008$ , one-tailed, against  $p = .002$ , one-tailed, in the original study. The effect size in this experiment ( $d = 0.28$ ), though quite healthy, was, however, smaller than that in Bem’s study ( $d = 0.42$ ), and this is in concordance with the prediction that minor changes in methodology would be reflected in altering the effect size (see introduction). Interestingly, in the first three weeks of experimentation and with 55 participants tested, the mean DR% score of 4.61 was almost identical to the mean score of 4.21 reported by Bem in Experiment 9, with 50 participants tested.

As in the original study, in the experimental condition of this study the “retrieval-induced forgetting” effect was observed: The absence of a significant condition effect on total recall indicated that in both experimental and control conditions participants recalled an approximately equal number of words, showing that practice exercises did not increase the total recall against the baseline shown in the control experiment. Rather, it appears that practicing with the selected set of words made their recall easier, while at the same time making recall of the not-practice words more difficult – the effect originally reported by Anderson, Bjork and Bjork [3].

However, the results of the control condition of this experiment differ from those reported by Bem. In the original experiment, the mean DR% score in the control condition was not significantly different from zero for the sample as a whole. In contrast, in this study the mean DR% score in the control condition ( $-5.21$ ) was highly significant, with the not-practice words being recalled significantly better than the

practice words. How could this unexpected result be interpreted?

One possible interpretation is that this result appeared due to the standard “interpersonal expectancy effect”, executed *via* subliminal messages that unintentionally slipped into the experimenter’s instruction to participants. Indeed, a summary of 345 studies of eight various categories (such as reaction time, animal learning, laboratory interviews, and other) has shown that when researchers expect certain results from their subjects they unwittingly treat them in such a way as to increase the probability that the subjects respond as expected [8]. The effect size varied across categories, but the reality of the phenomenon was beyond doubt.

However, this interpretation is unlikely: apart from the oral initial instruction, the rest of the procedure, as well as calculating the results scores, were strictly controlled by a computer, while the experimenter was separated from the participants by the opaque screen and maintained no voice contact with the participant. More importantly, the effect obtained in this study is an alternative to the interpersonal expectancy effect, as in this study the experimenter expected that in the control condition the mean DR% score should come to zero. The aforementioned considerations overrule the possibility to explain the results of this experiment by the regular interpersonal expectancy effect.

There is still, however, a possibility to interpret the results of this experiment as the experimenter expectancy effect, by assuming that the experimenter’s mind affected results directly, and not through the interpersonal expectancy bias. Further in this article, this assumption will be called “the non-standard experimenter effect”. Indeed, by expecting, in the experimental condition, practice words to be remembered better than not-practice words, the experimenter might “force” the RNG to pick the remembered words more frequently when selecting the practice words category sample. In that case, by default, the remaining not-practice words sample would contain a smaller number of remembered words than the number expected by a chance selection. Another way to influence the RNG would be inhibiting the se-



lection of non-remembered words while completing the practiced words sample, which would come up with the same results: the remembered words would be represented in the practiced words sample more frequently than they would be given a chance selection. As a result, the outcome of the experimental condition would look as though practicing with the practice words facilitated the recall of those words “backwards in time”. The advantage of this interpretation over the “retroactive facilitation of recall” interpretation is that it does not violate the common causality principle and puts the observed effect within the scope of scientific phenomena. A possible mechanism behind this kind of intentional distortion of random statistical processes by the experimenter has been proposed previously. According to “Decision Augmentation Theory” (DAT), the experimenter might access the sequence of random numbers by real-time clairvoyance, thus being able to determine the seeding of the system timer and, by doing so, to influence the target word selection [9,10].<sup>6</sup> The fact that in Bem’s program the random selection of practice words was implemented by Marsaglia’s pseudo random numbers generator (PRNG) algorithm, based on internal random function of REALbasic to seed the algorithm, makes involvement the aforementioned mechanism of the non-standard experimenter’s effect possible.

One possible objection to this hypothesis could be that, unlike a true RNG, PRNG is completely deterministic, and the experimenter has no control over the random numbers *via* a PK effect, like he or she might have if the true RNG were employed. This objection certainly points out to a limitation of the DAT. The only aspect of the PRNG that the experimenter can affect is the seeding of the already fixed sequence of numbers, by affecting the internal clock of the computer.

The main problem for the non-standard observer effect is to explain the results obtained in the control condition. How could it happen that, while in the control condition a zero mean DR% score was expected, the negative and significant DR% score was obtained, with not-practiced words being remembered better than practiced words?<sup>7</sup> One possibility to answer this question stems from the corollary that the non-standard observer effect explanation involves. According to this corollary, selection of practice and non-practice words by the PRNG is not affecting the participants’ recall “backwards in time”, but is a result of the experimenter’s direct interaction with the PRNG. If this is the case, then this selection becomes highly vulnerable to the experimenter’s subconscious biases and expectations. One of such subconscious biases is the subconscious attraction towards symmetrical patterns.

Indeed, according to gestalt psychology, our subconscious mind is biased by certain patterns, the so called “laws of gestalt” [11]. Among the laws of gestalt, the figure-ground distinction is particularly relevant. Indeed, when the experi-



Fig. (5). Asymmetrical and symmetrical figure-ground images.

mental and control conditions are separated, they are put in the figure/ground relationships to each other: the experimental condition was ran first, thus taking the role of the “figure”, whereas the control condition was run second, being therefore “a ground”. Most figure-ground patterns are asymmetrical, but some are symmetrical (see Fig. 5).

In biology, recent studies have found that humans and other animals are highly sensitive and attracted to symmetry in each other and often use it to gauge beauty and health during mate selection [12-14]. It appears that sensitivity to symmetry is ingrained into our behavior. Some scientists suggest that because our brains are so fine tuned to detect symmetry, it is possible that the tools that we use to determine the laws of nature and even our theories have symmetry in them partly because our brains like to latch onto the symmetric part of the universe [15]. But features that attract and are highly desirable can also be subconsciously projected onto reality. Because of such strong attraction towards symmetry, it is not inconceivable that, in some way, the symmetrical pattern was subconsciously projected onto results, even though consciously an asymmetrical pattern was expected. If this assumption is true, *then the effect observed in this experiment is a new version of the expectancy effect – the “subconscious expectancy effect”*. According to this version, *the experimenter’s subconscious image of the structure of this experiment (i.e., that the experimental and control conditions make a symmetrical figure-ground pattern) affected the results of the control condition, by shifting the DR% score to the opposite side from zero.*

One more effect observed was a negative correlation of the retroactive facilitation of recall with time of testing: as the testing progressed, the retroactive facilitation effect deteriorated. A similar deterioration of results had been observed in the earlier ESP study conducted by the author [16]. In that study a method designed by May [17] was used. Participants were encouraged to access information from the near future by drawing a picture of the photograph that they would be shown in about 15 minutes time. The experimenter then compared this picture to five photographs chosen randomly by computer out of the sample of 300, assigning each a rating between 0 and 100 to indicate the degree of similarity. After ratings for all five photographs had been entered into the computer and saved to disk, the computer randomly selected one of the five photographs as the target and this was displayed to the participant. As a measure of success to access the target picture a Z score was used.<sup>8</sup> As Fig. (6)

<sup>6</sup> The author thanks Adrian Ryan for pointing out to this explanation of the violation of random processes.

<sup>7</sup> It has to be noted, however, that in the control condition the “practice” words and “not-practice words” were only semantically different, due to the absence of practice. In contrast, in the experimental condition practice words differed from not-practice words also ontologically, owing to the fact that participants repeatedly saw and processed these words.

<sup>8</sup> Z score was calculated as (Target rating – Mean of trial ratings) / SD of trial ratings, where the target rating were the rating (from 0 to 100) assigned



**Fig. (6).** Mean Trail Z-score by time of testing, in Series 1 and 2 of Subbotsky & Ryan (2009) study.

shows, in Series 1 of this study, the Z score was at its highest in the first 2 weeks of testing, and deteriorated in the last two weeks of testing, producing a significant drop in results. In Series 2 of this study, which was an unsuccessful attempt to replicate a positive and significant Z-score obtained in one of the two conditions of Experiment 1, a similar drop in Z-score value with the time of testing was observed.

Interestingly, this effect is reminiscent of the “decline effect”, which was first noted by J.B.Rhine in his experi-

ments with Zener cards, and then reported in studies of a wide range of disciplines, from biology to parapsychology (Lehrer, 2010). The decline effect shows that statistically significant results of scientific experiments become less significant in subsequent replications, and the effect sizes shrink. The decline effect escapes simple explanations, such as regression to the mean, since its initial starting point is the effects, which are replicable, statistically solid and unlikely to be random deviations. Assuming that the non-standard observer effect is a real psychological effect that the experimenter’s mind directly makes on the experiment’s results, the experiment’s results could be influenced by a number of psychological and physiological factors, such as experi-

---

in the session to the target picture, and mean of trial ratings was the average of all 5 ratings assigned in that trial.



menter’s fatigue, drop in interest and motivation, bad mood, distraction from the experiment by extraneous circumstances, and so forth. Each of these factors, or their combination, might make the non-standard observer effect decline with the time of testing, both within one experiment, like in this study, and in subsequent replications of an initially successful experiment, like in Subbotsky & Ryan’s [16] study and in studies reviewed by Lehrer [18].<sup>9</sup>

If the symmetry interpretation is valid, the sequence of conditions is important. In particular, it can be expected that if the control condition runs first and the experimental condition second, then the zero mean DR% score, which is expected in the control condition, should come true, because, in the temporal succession of events, it becomes “a figure”, whereas the experimental condition becomes “the ground.”

This time, however, the experimental condition must conform to the subconscious expectation of symmetry, and should therefore produce the mean DR% score about the same as in the control condition, and on the opposite side from zero.

To test these expectations, Experiment 2 was run.

**EXPERIMENT 2.**

**Method**

*Participants*

Participants were graduates and undergraduates at Lancaster University (UK). Twenty-five participants (20 females and 5 males; age range 18 to 27 years, M = 20.39, SD =

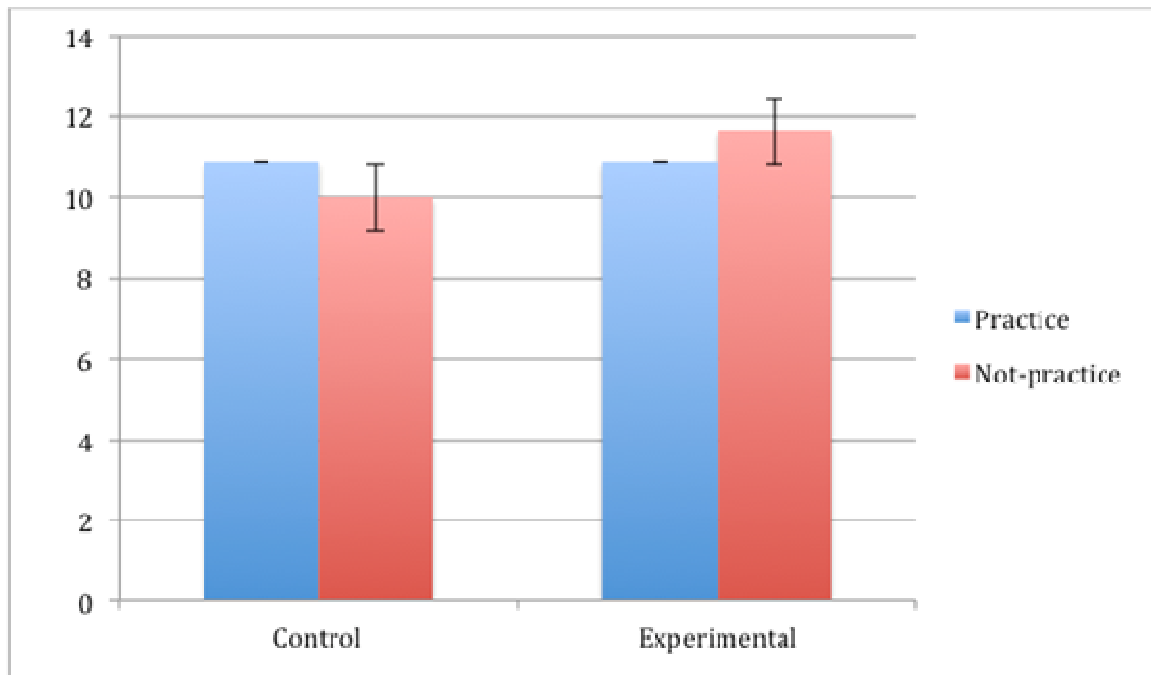
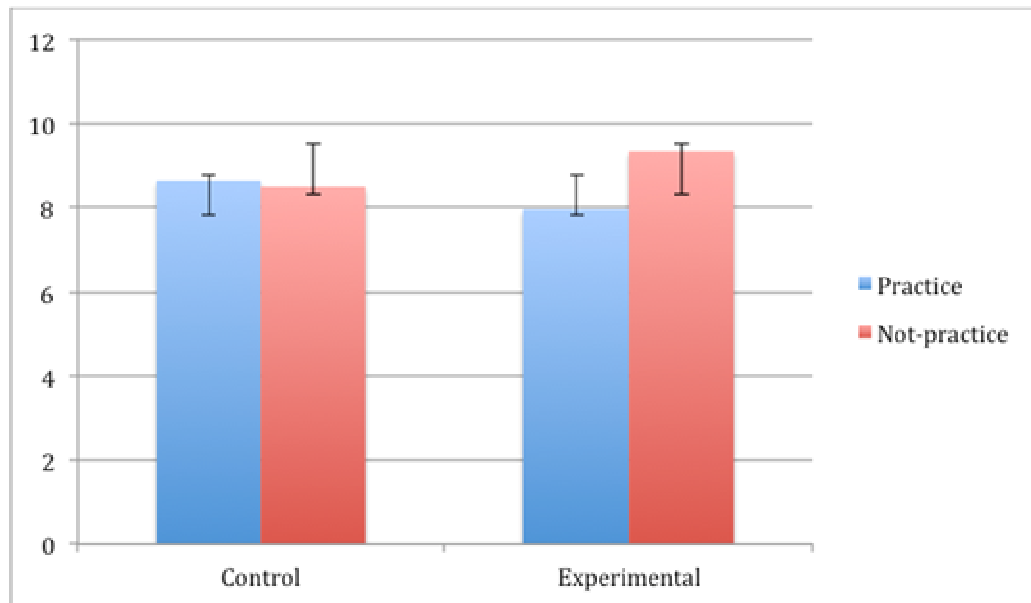


Fig. (7). Numbers of practice and not-practice words recalled in Experiment 2

Table 2. Words Recalled as a Function of Condition (Experimental and Control) and Type of Words (Practiced and Not-Practiced) in Experiment 2

	Control			Experimental		
	M	SD	%	M	SD	%
Practiced words	10.88	3.19	45.33	10.88	2.78	45.33
Not-practiced words	10.00	3.64	41.66	11.64	2.88	48.50
Total	20.88		42.17	22.52		39.83

<sup>9</sup> The non-standard observer effect’s sensitivity towards the aforementioned psychological and physiological factors may have implications on a sample size used in the experiment. It might be the case that increasing the number of participants in the experimental sample beyond a certain limit can mask fragile yet real non-standard experimenter’s effects, turning these effects into statistical artifacts. Thus, in the experimental condition of this study, increasing the number of participants from 50 to 75 decreased the significance level of the DR% score’s deviation from zero by 8.5 times, from 0.002 to 0.017, two-sided, with the effect size dropping from 0.46 to 0.28.



**Fig. (8).** Numbers of practice and not-practice words recalled in Experiment 3

**Table 3.** Words Recalled as a Function of Condition (Experimental and Control) and Type of Words (Practiced and Not-Practiced) in Experiment 3

	<u>Control</u>			<u>Experimental</u>		
	M	SD	%	M	SD	%
Practiced words	8.65	3.37	36.04	7.96	3.13	33.16
Not-practiced words	8.50	3.09	35.41	9.35	3.03	38.95
Total	17.15		35.72	17.31		36.06

2.51) contributed control sessions, and 25 (18 females and 8 males; age range 19 to 29 years,  $M = 20.4$ ,  $SD = 2.41$ ) contributed experimental sessions.

**Materials**

Materials were the same as in Experiment 1, retroactive facilitation of recall experiment. The application program was the one originally supplied by Daryl Bem.

**Design**

Design was the same as in Experiment 1, retroactive facilitation of recall experiment.

**Procedure**

Procedure was the same as in Experiment 1, retroactive facilitation of recall experiment, with one exception: the control session was run first, and the experimental session was run second.

**Results**

Like in Experiment 1, in this experiment the data were normally distributed. The mean DR% score in the control condition was 2.68%,  $SD = 15.76$ . The deviation of the mean DR% score from zero was not significant, one-sample  $t(24) = .851$ ,  $p = .40$ ,  $d = .17$ . In the experimental condition, the

mean DR% score was  $-3.06\%$ ,  $SD = 10.55$ . Again, the deviation of the mean DR% score from zero was not significant, one-sample  $t(24) = 1.51$ ,  $p = .14$ ,  $d = -.52$ . There was no significant difference between mean total number of words recalled in the experimental and control conditions,  $t(48) = -1.10$ ,  $p = .28$ .

Mean numbers (and SDs) of practice and not-practice words recalled as a function of condition are shown in Table 2, and illustrated in Fig. (7). A two-way ANOVA for condition (experimental and control) and type of words (practice and not-practice), with the repeated measures on the last factor, showed no main effects or interaction effects.

**Discussion**

The results of Experiment 2 supported the expectations based on the “projection of symmetry” explanation. In the control condition, participants’ DR% score was positive and not significantly different from zero. In the experimental session, the DR% score was negative, and of a similar effect size as it was in the control session. Phenomenologically, the patterns of practice and not-practice words recalled in the first- and second-run conditions in Experiment 2 are identical with those in Experiment 1.

Another expectation based on the “projection of symmetry” explanation of Experiment 1 and 2 results is that, if tem-

poral succession between control and experimental conditions is terminated, thus removing the figure-ground bond between conditions, the results in both conditions should become asymmetrical. This can be accomplished by interspersing trials of both conditions, but not in a random order, so that the experimenter still knows which condition is running.

One problem with Experiments 1 and 2 was that, as noted above, in the control condition the “practice” words and “not-practice words” were only semantically different, due to the absence of practice. In contrast, in the experimental condition practice words differed from not-practice words also ontologically, owing to the fact that participants repeatedly saw and processed these words. This partially undermines the “projection of symmetry” explanation, as long as what is meant by practice words in experimental and control conditions is not really comparable. In order to make practice words in the control condition of this experiment comparable with that in the experimental condition, practice words in the control condition should be made ontologically different from not-practice words, by giving the participants an opportunity to practice the practice words. In order to address this issue, Experiment 3 was conducted. If the projection of symmetry is determined by the figure-ground bond between the conditions in Experiments 1 and 2, then in Experiment 3, like in Bem’s original study, the results in both conditions should be different, with the experimental condition producing a significant and positive mean DR% score, and the control condition producing a near zero mean DR% score.

### EXPERIMENT 3.

#### Method

##### Participants

Participants were graduates and undergraduates at Lancaster University (UK). Twenty-six participants (14 females and 12 males; age range 18 to 28 years,  $M = 21.39$ ,  $SD = 2.31$ ) contributed experimental sessions, and 26 (8 females and 18 males; age range 18 to 31 years,  $M = 22.4$ ,  $SD = 3.41$ ) contributed control sessions.

##### Materials

Materials were the same as in Experiments 1 and 2, retroactive facilitation of recall experiment. For the experimental condition, the application program was the one originally supplied by Daryl Bem. For the control condition, the application was programmed by Adrian Ryan, in order to make the practice words flash subliminally, with each word appearing in the screen for the interval of 30 ms.

##### Design

Design was the same as in Experiment 1, retroactive facilitation of recall experiment.

##### Procedure and hypothesis

In order to eliminate the temporal factor in distinguishing between “a figure” and “the ground”, in this experiment, experimental and control sessions were divided into smaller sections, with approximately 5 participants per section, and the sections were interspersed in such a way that after each experimental section a control section followed. As a result,

the experimenter was aware which session (experimental or control) is running, however, none of the session preceded the other one, thus eliminating the possibility of the symmetry effect. It was expected that if the “symmetry” explanation of results in Experiments 1 and 2 is correct, then in this Experiment the results of the experimental and control sessions should be asymmetrical, with the control condition giving a near zero DR% score, and an experimental condition producing a mean DR% score significantly different from zero.

### RESULTS

Like in the previous experiments, in this experiment the data were normally distributed. The mean DR% score in the control condition was .76%,  $SD = 9.1$ . The deviation of the mean DR% score from zero was not significant, one-sample  $t(25) = .426$ ,  $p = .67$ ,  $d = .08$ . In the experimental condition, the mean DR% score was -4.06%,  $SD = 9.85$ . The deviation of the mean DR% score from zero was significant, one-sample  $t(25) = -2.1$ ,  $p = .046$ ,  $d = -.41$ . There was no significant difference between mean total number of words recalled in the experimental and control conditions,  $t(98) = .111$ ,  $p = .91$ .

Mean numbers (SDs) of practice and not-practice words recalled as a function of condition are shown in Table 3, and illustrated in Fig. (8).

A two-way ANOVA for condition (experimental and control) and type of words (practice and not-practice), with the repeated measures on the last factor, showed a significant interaction effect between the type of words and condition,  $F(1,50) = 4.30$ ,  $p = .043$ ,  $\eta^2 = .08$ . Simple effects indicated that in the experimental condition, not-practice words were recalled significantly better than practice words,  $t(25) = -2.61$ ,  $p = .015$ , whereas in the control condition the difference between the recall of the two types of words was not significant,  $t(25) = -.29$ ,  $p = .015$ ,  $p = .77$ .

### DISCUSSION

The results of Experiment 3, with correction, supported the notion that projection of symmetry in results between the two conditions is determined by the figure-ground bond between the conditions. As expected, when the figure-ground bond was removed, the DR% scores in experimental and control conditions were different: whereas in the control condition the effect size was .08 and the DR% score was not significantly different from zero, in the experimental condition the effect size was -.41 and the DR% score was significantly below zero, with not-practiced words being recalled significantly more frequently than practice words. Contrary to the expectation, however, the mean DR% score in the experimental condition was negative.

The puzzling question is why in this experiment, the expectation that the results in the experimental condition would be different from zero worked, but the value of the DR% score in this condition was opposite to one that had been expected. A possible answer is that the symmetry projection works across experiments, as well as within experiments. Specifically, since in Experiment 1, in the experimental condition the effect was significant and positive, then in Experiment 2 in this condition it was supposed to be significant and negative. However, due to a stronger figure-ground bond

with the control condition in Experiment 2, the effect in the experimental condition came close to zero. When, in Experiment 3, the figure-ground bond was removed, the effect in the experimental condition became significant and negative, thus maintaining the symmetry with the effect in the experimental condition of Experiment 1.

## GENERAL DISCUSSION

The data of the three experiments marginally support the assumption that the facilitation of recall effect reported by Bem is robust. At the same time, the data suggest a different interpretation of the nature of this effect than the interpretation offered by Bem. Looking at Pictures 3, 5, and 6, one can put this interpretation as follows.

If the retroactive facilitation of recall experiment is run with two conditions, one experimental and the other control, and the expected results in the conditions are asymmetrical, then in the experimental condition the expected retroactive facilitation of recall effect can show up, not show up, or show up in the reversed direction, with the not-practiced words being recalled better than the practiced words. When the conditions are run in temporal succession, the conditions' results tend to be symmetrical, and when the conditions are interspersed, the results tend to be asymmetrical. Even with the symmetry-asymmetry factor set aside, the invariable feature of the results of Experiments 1-3 was that these results oscillated, by taking 3 possible outcomes, with the mean DR% score being close to zero, or significantly different from zero to the positive or negative side.

Oscillating processes are known in science. One of such processes is a dissipative self-organizing system (see [19]). An example of this kind of system is the "chemical clock" -- the oscillating chemical reaction. Thus, in the Belousov-Zhabotinsky reaction chemical compound that includes bromine and acid, periodically changes color from red to blue and vice versa, briefly going through the phase of an unidentifiable color (a mixture of blue and red) in between, without any obvious intrusion of external factors (Zhabotinsky, [20, 21]). One theory behind this kind of reactions is the "oscillating system theory". According to this theory, in certain systems the energy-releasing reaction can follow at least two different pathways, and the reaction periodically switches from one pathway to another. The reaction is due to the symmetrical nature of chemical processes within these pathways: One of these pathways produces a specific intermediate, while another pathway consumes it. The switching of pathways is triggered by concentration of this intermediate. When the concentration of the intermediate is low, the reaction follows the producing pathway. When the concentration reaches a certain point, the reaction switches to the consuming pathway.

Accordingly, in the experimental condition of this study the system that included practiced and not-practiced words (the P-not-P system) behaved like an oscillating system: it either adopted the state of equilibrium (the "E" state), in which both kinds of words are recalled with the same frequency, producing a near zero DR% score (Experiment 2), or

dropped out of this basic state into two alternative states: the state in which practiced words are recalled better than not-practiced words (the "P prevail" state), producing a positive significant DR% score (Experiment 1) or the state in which not-practiced words are recalled better than practiced words (the "not-P prevail" state), resulting in a negative significant DR% score, (Experiment 3).

If the conditions of this study are counted as separate "experimental trials" involving the P-not-P system, then the empirically established probability of the P-not-P system to fall out of the "E" state is 50 %, since out of the 6 trials three (two in Experiment 2, and one in Experiment 3) produced near zero DR% scores, and the other three (two in Experiment 1, and one in Experiment 3) produced DR% scores significantly different from zero. Interestingly, in the control condition of Experiment 1, the P-not-P system distinguished between the two classes of words (symbolically named as "practiced" and "not-practiced" ones) by falling into the not-P prevail state even though these classes of words were ontologically identical. This effect is unusual and runs contrary to the one consciously expected by the experimenter, since, with the absence of practice, participants should have recalled practice and not-practice words with approximately the same frequency.

One possibility to make this effect meaningful is to speculate that the experimenter's mind, which is evolutionary biased towards symmetrical patterns, subconsciously projected this pattern onto results, forcing the control condition to produce results symmetrical to those in the experimental condition, even though consciously a zero DR% score was expected in the control condition. If such assumption is made, then it becomes possible that this subconscious expectation *brought the P-not-P system out of the "E" state in the control condition of Experiment 1, the (non-standard observer effect)*.

The symmetry between results of experimental and control conditions was observed when the conditions were run in succession (Experiments 1 and 2 of this study), and was not observed when conditions were interspersed (Bem's original experiment, and Experiment 3 of this study). This fact adds to the suggestion that the results were shaped by the non-standard observer effect: the conditions produced symmetrical results only when they were related to each other as figure and ground -- the feature which is determined by the nature of human perception. The results also suggest that the symmetry bias within experiments supersedes the one between experiments. Thus, according to the symmetry bias, the results in the experimental condition of Experiment 2 should have been negative and significant, to maintain the symmetrical pattern with the results of the experimental condition of Experiment 1. Instead, in Experiment 2 the results scored near zero, to maintain the figure-ground bond with the control condition of this experiment, which was run first. When the figure-ground bond between conditions was removed in Experiment 3, the results of the experimental condition scored significantly below zero, thus reinstating the symmetrical pattern with the matching condition of Experiment 1.

To summarize, the study's overall results suggest that practicing with the practice words is unlikely to be the factor that disturbs the "E" state, like it was assumed in Bem's original paper. Rather, one of the particular states that the P-not-P system adopts as a result of the disturbance of the "E"

state might look as if the participants' recall benefited from subsequent practice with the practice words.<sup>10</sup> At the same time, this study suggests that Bem's experimental methodology is a sensitive device that can react to the experimenter's mind, i.e. to the experimenter's subconscious mental image of what is expected to happen. When Bem's methodology was applied, the distribution of practice and not-practice words significantly deviated from the one predicted by common logic (i.e., from the 50-50 proportion), in three out of six experimental conditions. Assuming that each condition was run independently from other conditions, the selection of P and not-P words conformed to binomial distribution, and given the 5% confidence interval (two-sided), the probability of this to happen by chance is 0.002. It is tempting to assume that the mechanism underlying these deviations is the effect of experimenter's subconscious expectations on functioning of the PRNG, on which the frequencies of practice and not-practice words are based. A theoretical possibility of such interaction between the mind and the PRNG results have been shown previously [10,11]. If the DR% score in the initial condition deviates significantly from zero, then in subsequent follow up experiments the DR% score tends to change, conforming to the symmetrical oscillating pattern.

Needless to say that some assumptions made in this paper are of strictly preliminary nature and on the margins with speculations. Nevertheless, in the domain where the body of data is scarce and established theories nonexistent, making hypothetical assumptions is the only way of apprehending unpredictable and sometimes mind-bending results.

#### CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

#### ACKNOWLEDGMENT

The authors thanks Adrian Ryan for his assistance with programming and writing up Experiment 1, and Jessica Utts and Damon Berridge for their assistance on some statistical issues.

#### REFERENCES

- [1] Bem, D. Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *J Pers Soc Psychol* 2011; 100: 207-25.
- [2] Bjork RA. Repetition and rehearsal mechanisms in models of short-term memory. In: Norman DA, Ed. *Models of human memory*. New York: Academic Press 1970; pp. 307-30.
- [3] Anderson MC, Bjork RA, Bjork EL. Remembering can cause forgetting: retrieval dynamics in long-term memory. *J Exp Psychol Learn Mem Cogn* 1994; 20: 1063-87.
- [4] Galak J, LeBoeuf, RA, Nelson LD, *et al*. Correcting the past: Failures to replicate *Psi*. 2012.
- [5] Ritchie SJ, Wiseman R, French C. Failing the future: Three unsuccessful attempts to replicate Bem's "Retroactive Facilitation of Recall" effect. *PLoS one* 2012; 7(3): 1-5.
- [6] Mathews RC, Tulving E. Effects of three types of repetition on cued and noncued recall of words. *J Verb Learn Verb Behav* 1973; 12: 707-21.
- [7] Francis W, Kucera H. *Frequency analysis of english usage: lexicon and grammar*. Boston: Houghton Mifflin 1982.
- [8] Rosenthal R, Rubin DB. Interpersonal expectancy effects: the first 345 studies. *Behav Brain Sci* 1978; 3: 377-15.
- [9] May EC, Utts JM, Spottiswoode SJP. Decision augmentation theory: Applications to the random number generator database. *J Sci Explor* 1995; 9(4): 453-88.
- [10] May EC, Utts JM, Spottiswoode SJP. Decision augmentation theory: Toward a model for anomalous mental phenomena. *J Parapsych* 1995; 59: 195-20.
- [11] Sternberg R. *Cognitive Psychology*. New York: Thomson Wadsworth 2003.
- [12] Chen CC, Wu JH, Wu CC. Reduction of image complexity explains aesthetic preference for symmetry. *Sym* 2011; 3: 443-56.
- [13] Fink B, Neave N, Manning JT, *et al*. Facial symmetry and the 'big-five' personality factors. *Person Ind Diff* 2005; 39(3): 523-29.
- [14] Rhodes G, Zebrowitz LA. *Facial attractiveness - evolutionary, cognitive, and social perspectives*. New York: Ablex 2002.
- [15] Livio M. *The equation that couldn't be solved: how mathematical genius discovered the language of symmetry*. New York: Simon and Shuster 2006.
- [16] Subbotsky E, Ryan A. Motivation and belief in the paranormal in a remote viewing task. 2009. Available from: [www.lancs.ac.uk/~staff/subbotsk/](http://www.lancs.ac.uk/~staff/subbotsk/)
- [17] May EC. *Anomalous Cognition: Two Protocols for Data Collection and Analysis*. 2006. Available from: <http://www.lfr.org/LFR/-csl/library/ACtrials%26Analysis.pdf>
- [18] Lehrer J. *The Truth Wears Off*. *The New Yorker* 2010; 52-57.
- [19] Prigogine I, Stengers I. *Order out of Chaos: Man's new dialogue with nature*. New York: Flamingo 1984.
- [20] Zhabotinsky AM. Periodical process of oxidation of malonic acid in a solution. *Biophysics* 1964; 9: 306-11.
- [21] Zhabotinsky AM. Concentration caused oscillations. *Moscow: Nauka* 1974.

Received: June 21, 2013

Revised: September 27, 2013

Accepted: September 29, 2013

© Eugene Subbotsky; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.

<sup>10</sup>To some extent, it is a matter of luck whether the significant deviation of the DR% score from zero happens or doesn't happen in the first attempted experiment. If the first experiment fails to bring the P-not- P system out of the "E" state, thus supporting the null hypothesis, then the follow up experiments are unlikely to be attempted by the same experimenter. If we are to comprehend the reality of non-standard psychological effects, we may have to part with the model of mechanical replicability of data and take on a probabilistic model instead.