

Examining Retroactive Facilitation of Recall:
An Adapted Replication of Bem (2011, Study 9) and Galak and Nelson (2010)

Steve Snodgrass

President, Omaha UFO Study Group

First Posted Online: September 27, 2011

Address correspondence to admin@omahaufostudygroup.com

Abstract

The author performed a procedure to test precognitive features of word recall termed “Retroactive Facilitation of Recall” in Bem (2011) by adapting the failed replication procedure performed by Galak and Nelson (2010.) Despite enhancing subjects’ opportunities to demonstrate the effect using handwritten practice scenarios, this small group of UFO and paranormal enthusiasts failed to replicate the phenomenon as well. This paper provides a brief description of similarities and differences between Bem’s study #8 and #9, and a description of how Galak and Nelson’s procedures were adapted in this study. We conducted (1) an informal demonstration of the effect during an on-air radio broadcast with a single in-studio subject and (2) an experimental procedure using volunteer subjects in a joint meeting of the Omaha UFO Study Group, University of Nebraska at Omaha UFO Study Group, and chapter meeting of Nebraska MUFON, who agreed to collective administration of Galak and Nelson’s online interactive website application.

Examining Retroactive Facilitation of Recall:
An Adapted Replication of Bem (2011, Study 9) and Galak and Nelson (2010)

Bem (2011) published a thorough series of experimental procedures demonstrating convincing evidence supporting the psi notion of precognition. Bem's experimental evidence for retroactive influences from the future on cognition and affect received media attention as well as penetrating attention from skeptics. Soon after Bem's publication in the prestigious peer-reviewed *Journal of Personality and Social Psychology* skeptic Richard Wiseman sought to organize a resource for those interested in conducting replication studies to verify Bem's groundbreaking work. Wiseman also challenged Bem's scoring methodology.

One replication attempt is Galak and Nelson (2010) whose experiment failed to replicate Bem's study #8, Retroactive Facilitation of Recall I. Instead of replicating the study in a laboratory, Galak and Nelson chose to develop a guided interactive online slide presentation based on Bem's design. They performed the replication by developing an online participant pool. Since publishing their initial findings in an online article, Galak and Nelson's study continues, sort of, offering this same web-based interactive program to an online audience free of charge via an internet website. Ongoing results are fed back to participants in the form of their "ESP score" at the conclusion of their participation in the study. The ongoing site has amassed some 1,400 participants as of the date of this writing, with ongoing results that remain supportive of a null hypothesis.

They reach two conclusions: "we conducted a very close replication of Bem (2011, Study 8) and failed to obtain a reliable result" and "we have not disproven Bem (2011)...but are merely trying to add more data relevant to the question." In endnotes to

their study Galak and Nelson report they led student study groups at their respective universities and offered odds at replicating Bem's results successfully at 4:1 against and 19:1 against.

In contrast to the authors from Carnegie Mellon University and the University of California, Berkeley, one might assume by the authorship of this paper that our subject pool, comprised of nine UFO enthusiasts, might involve experimenters who would place the odds of successful replication much higher. This would be a correct assumption. Although a polling of odds did not take place, the author of this paper estimated the odds of successful replication of this phenomenon at 1:1. Should this immediately raise a flag that our research must be fraught with fraud and experimenter bias? Perhaps. If, however, our group is consistent with what it claims as their mission, to look for logical and rational explanations of the UFO phenomenon through the lens of science, then only the reader can reach a conclusion as to the extent to which experimenter bias played a role in our experiments.

Richard Wiseman's challenge to Bem's methodology involved the scoring of the output of participants, i.e. how do we know your student volunteers did not entertain bias in the manner in which they scored participants' word recall lists? Bem has responded to this adequately. Galak and Nelson seem to agree. However, the latter chose to "remove even that small human element from the analysis." Galak and Nelson decided to allow a computer to score results in such a way that a participant who recalls the word "Apple" but types it as "Aple" would be excluded. In our study, the words "Spaghetti" and "Lasagna" (both control words) were often misspelled, but were included along with

every other word from participants' recall list. (The word "Apple" was not incorrectly spelled in our study).

Galak and Nelson's conclusion, "it would be preferable to have a procedure which removes even that small human element from the analysis" is sound if their premise is true. Their premise is "given the history of Psi research is marked by subtle influences of experimenter bias." The author of this paper disagrees with this premise and believes a more accurate reflection of Psi research is that it is "marked by *not unreasonable, but exaggerated scrutiny regarding* subtle influences of experimenter bias." While Galak and Nelson's placing odds against successful replication is informative, this author believes the inclusion of "casting lots" in this way suggests greater potential for greater intrusion of experimenter bias. Galak and Nelson's stated goal to "clarify the effect" may have driven their choice of which of Bem's experiments they chose to replicate and which scoring procedures would be allowed.

As the methodology of our experiment describes, we chose to offer the greatest possible chance at replicating Bem's results by adapting the above studies in two ways. First, this study allowed a single individual, the author, to score the results. "Trust no one" and "I know what I saw" are themes in UFOlogy. In this case, trust must be placed in the author that words spelled like "lasania" are straightforwardly decipherable. Second, in order to give participants every opportunity to replicate Bem's result successfully, we modified the practice session of the experiment for the following reasons.

Bem performed an additional experiment (#9) or "Retroactive Facilitation of Recall II" wherein participants were provided additional word practice opportunities

following the recall procedure. Bem took a further step to ascertain the nature of retroactive facilitation by simply allowing participants to see separate word lists in each category (clothing, food, occupations, animals.) Why? Because once the initial word list is recalled and recorded, it should not matter how much practice participants are provided. Any amount of additional practice of target words might only strengthen any observable effect of retroactive facilitation. The reader will note from Bem (2011) that p values for Study #8 were 0.029 whereas Study #9 p values were 0.002. It was our belief that if we were unable to replicate Bem's study precisely, we should follow Bem's footsteps by either following the protocol in Study #9, or venture farther by enhancing the practice procedure of target words, thus "clarifying the effect." This is what our experiments sought to accomplish.

Study #1: Informal Demonstration

An informal demonstration of the principles of Bem's study was explored with a single subject on a live web streaming radio broadcast of the *Spooky Action at a Distance Radio Show*. (See September 9, 2011 episode available in audio archive at www.omahaufostudygroup.com.) The host and producer of the show had expressed interest in the phenomenon of precognition and Bem's work. Consequently, the author, a guest on the broadcast, discussed the phenomenon of retroactive facilitation of recall. The author was requested, and consented, to attempt to perform a rough demonstration of the phenomenon live in the studio.

Method

The subject, a female in her 20s, was informed that the experiment involved a memory recall test and consented to volunteer. A list of 16 nouns developed by the

author were written down and verbally spoken aloud to the subject, who was then asked to study this list for approximately 30 seconds. She was then asked to recall as many words as she could and write them down. Approximately one minute transpired before the subject indicated she could not remember any additional words. The recall list completed by the subject was then removed from view. Following this procedure the subject was provided a list of four words from the initial word list (odd words #2, #4, #6 and #8) selected by the experimenter as target words before the study began. These four words were written down on a sheet of paper as prompts. The subject was asked to hand write each of these practice words six times. The author tabulated the number of control words (not practiced afterward) and practice words (rehearsed following the recall) to ascertain the likelihood that retroactive facilitation of recall (precognition) had an effect on words recalled.

Results

The subject recalled eleven of sixteen words (68%) in total: the first six words, the last 3 words, and two additional words from the initial list presented to her. Two of the words recalled were practice words (out of four possible, 50%) and nine words recalled were control words (56% of the total.) The last word recalled by the subject, “airplane”, was written with a question mark. A statistical analysis consistent with Bem and Galak and Nelson’s studies could not be performed because only four words of sixteen (25%) were chosen as target words to be recalled. Bem’s study and Galak and Nelson’s study used 48 words in total, 50% of which (24) were randomly chosen as words to be practiced.

Discussion

The subject explained the method she used for generating her recall list. She explained the first six words she recalled were due to the use of a mnemonic device, creating a sentence based on the first letter of words that were presented to her. The last three words on the list, she explained, remained in her short-term memory as she attempted to reconstruct the list beginning with the last word. The subject explained one additional word had remained in visual short-term memory, as an image of the written word itself. After exhausting her mnemonic devices and short-term memory, she paused and then arrived at the word “airplane”, placing a question mark behind it.

Among the many variables that interfered with any possible interpretation that retroactive facilitation of recall could be measured with this methodology are: the length of the word list (only 16), the percentage of target to control words (25% instead of 50%) and the instruction given to the subject (that this would be a test of memory) and lack of randomization procedure. The result of this experiment demonstrated the power of mnemonic devices and confirmed well-known features of short- and long-term memory. However, it was interesting to note the subject’s description of how the last word on her recall list, “airplane” differed from the others. “I couldn’t remember any more words, and then “airplane” popped into my mind at the end, so I wrote it down.” This does not confirm nor deny the potential that retroactive facilitation was present, but simply suggests another avenue of study. Future studies may choose to focus on the hypothesis that words toward the end of a subjects’ recall list, after they have exhausted any other methods of recall, may be more influenced by future practice of those words. This is a feature we considered in our less informal experiment, Study #2.

Study #2: Collective Experimental Procedure

The purpose of this study was to more closely approximate Bem's experiment by using Galak and Nelson's online computerized application as the stimulus. A second purpose of this study was to preserve one aspect of Study #1; namely, providing subjects a more lengthy practice session involving handwritten practice of each target word six times. Our hypothesis is consistent with Bem: retroactive facilitation of recall will be more pronounced given greater opportunity for practice of target words. Our second hypothesis is that the last word in each subject's recall list will more than likely be a target word (a word the subject will subsequently practice.)

Method

Nine subjects (eight men and 1 woman) at a regular meeting of the Omaha UFO Study Group and the University of Nebraska UFO Study Group agreed to participate in an "ESP experiment." A classroom in the Durham Science Center building on the UNO campus allowed Galak and Nelson's website application to be displayed via video projector. The author operated the computer mouse, guided participants through the application, and provided each participant a pen/pencil and two pre-printed handouts. Handout #1 asked for the subject's name, age, gender as well as a space to circle their answers to the two stimulus seeking questions. It also served as the subject's recall sheet. Handout #2 provided a space for the subject's demographic information (for matching purposes) and contained a total of 144 lines on which to practice target words generated by Galak and Nelson's application during the rehearsal procedure. The author entered the required information to allow the application to proceed through the recall portion. The author entered no information on the recall screen, but instead asked participants to

write on their Handout #1 as many of the words they could recall. During the practice session the author asked the subjects to call out each word to be clicked and highlighted in red. The author manually typed each word to allow the application to proceed. At each category break, subjects were asked to handwrite each target word a total of six times on their Handout #2. Results were hand scored by the author.

Results

Not all subjects chose to answer the age and gender questions, nor the stimulus seeking questions. Consequently, data comparing stimulus-seeking questions in a separate analysis was not available. Control and practice words were hand scored by the author and entered into a free one-sample t-test calculator (www.graphpad.com). The total sample was a mean DR% of -0.021%, $t(8)=1.0067$, $p=0.3436$. In scoring the last word on each subject's recall list, a total of 5 were control words and 4 were practice words, well within odds of chance.

	N	P	C	Mean (DR%)	Statistic
Galak/Nelson	112	8.09	8.43	-1.35%	$t(111) = 1.31, p = .194$
Snodgrass	9	5.67	6.78	-0.021%	$t(8) = 1.01, p = 0.3436$

Discussion

Recall the difference between Bem's Study#8 and Study #9. The latter involved an extra step in the practice session. Subjects were informed the 24 practice words were categorized as "animals", "occupations", "food" and "clothing" and also provided a list of all four categories of words separately. The purpose, according to Bem, was to enhance subjects' ability to demonstrate the phenomenon. Results from Bem's study #8

for the total sample was a mean DR% of 2.27%, $t(99) = 1.92$, $p = .029$, $d = 0.19$ compared to study #9 where the total sample was a mean DR% of 4.21%, $t(49) = 2.96$, $p = .002$, $d = 0.42$. Rather than replicating the weaker of Bem's experiments with less precision, we chose to follow Bem's logic and even further enhance subjects' opportunity to demonstrate precognitive effects. In our study the null hypothesis is supported.

We believe the number of subjects involved in our study (nine UFO enthusiasts), was a representative sample for the following reasons. In a metropolitan community with an approximate population of 500,000 about half may believe in some type of paranormal phenomena. However, significantly fewer choose to actively engage in discussion about UFOs, alien abductions, ESP, paranormal phenomenon and cryptoscience. While a much larger population exists who purportedly believe and accept these phenomena as truly unexplained and possibly real, the number of individuals prepared to openly discuss such phenomena remains markedly small. The reader is free to speculate on why this is the case. We believe our sample is representative of a population of individuals who agree with Carl Sagan's maxim "Extraordinary claims require extraordinary evidence." yet have not succumbed to a prevailing reductionist philosophy in contemporary science that promises an unfalsifiable physicalist or materialist explanation of the universe as promulgated by pseudoskeptics. Our sample represents a body of individuals who believe "ordinary claims deserve equal suspicion."

It is possible we failed to replicate Bem's results for the same reasons Galak and Nelson cite: greater heterogeneity in their sample, or quantum effects related to the features of Bem's experimental room that somehow prevented precognitive effects. Perhaps among the more mundane explanations not mentioned by Galak and Nelson is

how the sheer power of Bem's mind and credibility as a social scientist telekinetically distributed the words in Galak and Nelson's study so that those easiest to spell (and present in the book *Dick and Jane*) would disperse evenly, thus increasing the likelihood that their study would fail to replicate his. Such an incredible ability could explain a conspiracy among the editors of the *Journal of Personality and Social Psychology* in choosing not to publish replication studies that do not support Bem's findings; a allegation posited by Wiseman.

Conclusions

Despite our enhanced practice session, causing fatigue and discomfort on the part of subjects asked to handwrite 144 words on a single sheet of paper, we failed to replicate the same results as Bem using Galak and Nelson's automated procedure. We achieved approximately the same results as Galak and Nelson and did not observe any effect of retroactive facilitation of recall. Interestingly the DR% mean in our study was closer to zero. This does not lend support to either precognition, a lack thereof, or what Charles Tart refers to as "psi-missing". Psi-missing might be evident if any replication of Bem's study came in with a DR% at -2.00% or less at a level of statistical significance, meaning subjects somehow selected control words more than practice words at odds against chance. It would be suggestive of the sheep/goat effect as described by Schmeidler and Murphy (1946). Galak and Nelson's DR% came in at -1.35%.

However, despite full scoring opportunity of all subject's responses, regardless of spelling errors, our study supports Bem's findings that hand scoring of subject's responses did not introduce experimenter bias. This finding adds further confirmation

that there is no foundation to Richard Wiseman's criticism of Bem's scoring procedures.

It bears repeating in a different way. By using researchers and subjects completely open to ridicule and devastating experimenter bias, we chose to ignore the potential for such derision by following the logic of Bem's hypotheses. If anyone is to be charged with bias, it ought to be our group. Yet, here are our results. We cannot confirm with our data that retroactive facilitation of recall (precognition) in Bem's study was replicated successfully here. We cannot reject the null hypothesis. We UFOlogists follow the scientific method just as well, and we reach our final conclusion from a position of complete open-mindedness unencumbered by potential experimenter biases arising from stances of reductionism and pseudoskepticism. Finally, and alternatively, our group notes that time-travel experiences reported by extraterrestrial contactees are also neither supported nor rejected as a result of our study. So our conclusion is simply: ..."Shit."

References

- Bem, Daryl J. (2011) Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect. *Journal of Personality and Social Psychology*, 100, 407-425.
- Galak, Jeff and Nelson, Leif D. (2010) A Replication of the Procedures from Bem (2010, Study 8) and a Failure to Replicate the Same Results. Website publication available at <http://ssrn.com/abstract=1699970>
- Tart, Charles T. (2009) *The End of Materialism: How Evidence of the Paranormal is Bringing Science and Spirit Together*. Oakland, CA: New Harbinger Publications, Inc.
- Schmeidler, G.R., & Murphy, G. (1946). The Influence of Belief and Disbelief in ESP upon Individual Scoring Level. *Journal of Experimental Psychology*, 36, 271-276.