

A Summary of “Feeling the Future: A Meta-analysis of 90 Experiments on the Anomalous Anticipation of Random Future Events” by Bem, Tressoldi, Rabeyron & Duggan.

This document is a brief, non-technical summary of the article “Feeling the Future: A Meta-analysis of 90 Experiments on the Anomalous Anticipation of Random Future Events” by Bem, Tressoldi, Rabeyron, & Duggan (2015). The article itself can be viewed, downloaded, and shared without charge from <http://f1000research.com/articles/4-1188>.

In 2011, the *Journal of Personality and Social Psychology* published a report of nine experiments by Cornell Professor Daryl Bem purporting to demonstrate that an individual’s cognitive and emotional responses can be influenced by randomly selected stimulus events that do not occur until after his or her responses have already been made, a generalized form of the phenomenon traditionally denoted by the term *precognition* (Bem, 2011).

Each of the experiments modified a well-established psychological phenomenon by reversing the usual time-sequence of stimulus-response events so that an individual’s responses were obtained before the putatively causal stimulus events had occurred. The hypothesis in each case was that the time-reversed or precognitive version of the experiment would produce the same result as the standard non-time-reversed experiment.

For example, one of psychology’s oldest and well-established phenomena is that individuals are more likely to make a response that had been rewarded in the past than one that had not been rewarded. The time-reversed version of this phenomenon tested whether individuals are more likely to make a response that would be rewarded in the near future.

On each trial of the experiment, the participant was presented with two curtains displayed side-by-side on a computer screen. The participant was told that an erotic photograph was behind one of the curtains and a blank wall was behind the other. The participant’s challenge was to select the curtain that concealed the erotic photograph. Actually however, the computer waited until the participant had already made his or her choice before it randomly selected the curtain that would conceal the erotic photograph. If the participant had selected that curtain, it opened to reveal an erotic photograph and the trial was scored as a hit; if the participant had selected the other curtain, it opened to reveal a blank gray wall and the trial was scored as a miss. A participant’s final score was the percentage of hits achieved. This experiment was titled “Precognitive detection of erotic stimuli.” Several other well-established psychological phenomena were tested using the time-reversed procedure; all showed the predicted precognitive effects.

The controversial nature of these results prompted the Journal’s editors to publish an accompanying editorial justifying their decision to publish the report and expressing their hope that attempts at replication would follow. Most scientists agree that the critical test of controversial findings is whether or not independent investigators can successfully replicate them, and the major analytic tool for answering this question is called meta-analysis: Whereas the analysis of a single experiment summarizes and evaluates observations across trials or participants, a *meta*-analysis summarizes and evaluates results across experiments.

Several years before the formal publication of the 2011 article, Bem began to encourage such replications by offering free, comprehensive packages that included detailed instructions for conducting the experiments, computer software for running the experimental sessions, and programs for collecting and analyzing the data. As a result, two years after the publication of Bem’s experiments, we were able to locate 69 independent replications of those experiments and a few related precognition experiments that were not designed to be replications of those experiments. When Bem’s own experiments are included, the complete database comprises 90 experiments from 33 different laboratories located in 14 different countries. A total of 12,406 individuals participated in these experiments.

Statistical Analysis of the Results

In evaluating the results of an experiment or set of experiments, two quantities are of prime interest: Its “effect size”: How big is the observed effect? And its “statistical significance,” the probability that the observed effect might simply be due to chance.

Effect Size. Expressing the effect size achieved by an experiment is sometimes quite straightforward. For example, in the erotic detection experiment described above, a participant chooses between two equally-likely curtains on each trial. If there is no precognition operating in the experiment—if only chance is operating—than we would expect participants to achieve an average hit rate of 50%. The hit rate actually observed in Bem’s original experiment was 53%.

This effect size may appear trivially small, but it is not. For example, the United States Presidential election of 2008 was considered to be a near-landslide victory for Barack Obama because he won 53% of the popular vote. Another example is the roulette wheel. It contains 36 numbered holes on which a player can place a bet. In addition to betting on a single hole, a player can bet that the ball will land on an even- or an odd-numbered hole or on a red- or black-colored hole. To the player,

these might appear to be even 50-50 bets. But an American roulette wheel actually contains two additional holes, “0” and “00.” If the ball lands on either of these, the casino wins. This means that casinos actually win 53% of these bets—and they’re not complaining. In principle, a roulette player with the same degree of precognitive ability as participants in the erotic-detection experiment could erase the casino’s advantage. European roulette wheels contain only one additional hole, so those casinos win only 51% of the bets.

Because different experiments measure many different kinds of variables, psychologists have developed standard measures of an experiment’s effect size that are independent of whatever variable was actually measured. As a rough rule of thumb, a standardized effect size of .8 or greater is considered to be “large.” An example is the obvious average height difference between 13- and 18-year old girls. An effect size of .5 is considered to be “medium,” which is still big enough to be visible to the naked eye of someone with experience observing the variable; an example is the average IQ difference between clerical and semiskilled workers. Finally, an effect size of .2 is considered to be “small,” and is typical of effect sizes in many areas of psychological research (Cohen, 1988). For example, the average effect size of 25,000 social psychological experiments spanning 100 years of research is .21 (Richard, Bond, & Stokes-Zoota, 2003). The 53% result of Bem’s erotic-detection experiment translates into a standardized effect size of .25, and the average effect size across all nine of his experiments is .22.

A new method specifically designed for estimating the “true” effect size of experiments in a meta-analysis has recently been developed and tested extensively on pre-existing data (Simonsohn, Nelson, & Simmons, 2014). Using this method, the overall effect size of our database is .20, very similar to that of Bem’s original experiments. (The older, more traditional method for estimating the effect size yields a smaller estimate of approximately .10.) If we exclude Bem’s original experiments, then the effect size of the 69 independent replications in our database is .24.

Again it is instructive to compare these effect sizes with others from publicly familiar examples. An example is the widely publicized medical study that sought to determine whether a daily dose of aspirin can prevent heart attacks (Steering Committee of the Physicians Health Study Research Group, 1988). That study was discontinued after six years because it was already clear that the aspirin treatment was effective and it was considered unethical to keep the control group on placebo medication. Even though the study was considered a major medical breakthrough, the size of the

aspirin effect is only about .07, approximately one third the size of the precognition studies (McCartney & Rosenthal, 2000).

Statistical Significance. Psychologists have adopted the convention that an effect may be called “statistically significant” if the probability that it would have occurred by chance is less than 1/20 (or 5% or .05). Survey researchers use this same convention: If a survey researcher announces that candidate A is ahead of candidate B, it means that the difference between the two of them is sufficiently large that it would occur less than 5% of the time if only chance were operating. If the difference between the two candidates is too small to satisfy this criterion, they are said to be in a statistical tie or dead heat.

The probability that the results of Bem’s original experiments are due to chance is 10^{-11} , much smaller (i.e., better) than the .05 criterion for statistical significance. In other words, the probability that his results would have occurred by chance is approximately 1 in 100 billion. The significance level for the 69 independent replications of his original experiments in our database is approximately 10^{-5} or 1 in 100,000, again much smaller than the .05 criterion for statistical significance.

The results of our meta-analysis do not stand alone. Bem’s precognitive experiments can be viewed as conceptual replications of what are known as “presentiment” experiments, in which physiological measures of participants’ emotional arousal are monitored as they view a series of pictures on a computer screen. Most of the pictures are emotionally neutral, but on randomly selected trials, a highly arousing erotic or negative image is displayed. As expected, participants show strong physiological arousal when these images appear, but the important “presentiment” finding is that the arousal is observed to occur a few seconds before the picture actually appears on the screen—even before the computer has randomly selected the picture to be displayed. In a meta-analysis of presentiment experiments, the effect size was .21, virtually identical to both Bem’s experiments and those in our meta-analysis (Mossbridge, Tressoldi, & Utts, 2012).

The Problem of Missing Studies: The File-Drawer Effect

It is widely acknowledged that successful studies in scientific fields are more likely to be submitted and accepted for publication than unsuccessful studies. As a consequence, conclusions that are drawn from meta-analyses based on the known studies can be misleading because we don’t know how many unsuccessful studies are left languishing in the file drawers of their investigators—hence

the term File-Drawer Effect. For our meta-analysis, we expended intensive effort to identify and include both published and unpublished replication attempts. There are also several statistical techniques for estimating the extent to which the absence of unknown studies might be biasing a meta-analysis. In our article we report on nine of these techniques.

The most commonly used technique for estimating how many unsuccessful studies are likely to be missing from a database examines the relationship between the effect size of each study and its corresponding number of sessions or participants. This technique estimates that there are likely to be only eight unsuccessful studies missing from our database. We also calculated the number of unsuccessful studies that would be required to nullify the overall effect size of our database if they existed and were to be included. The answer was 544 unsuccessful studies. That is, there would have to be 544 unsuccessful studies missing from our database to reduce its overall effect size to a trivial level. In conjunction with the results from all the other analyses, we therefore concluded that the file-drawer effect has not compromised our meta-analysis.

General Discussion

Precognition is one of several phenomena in which individuals appear to have access to “nonlocal” information, that is, to information that would not normally be available to them through any currently known physical or biological process. These phenomena, collectively referred to as *psi*, include *telepathy*, access to another person’s thoughts without the mediation of any known channel of sensory communication; *clairvoyance*, the apparent perception of objects or events that do not provide a stimulus to the known senses; and *precognition*, the anticipation of future events that could not otherwise be anticipated through any known inferential process.

Psi is a controversial subject, and most academic psychologists do not believe that psi phenomena are likely to exist. A survey of 1,188 college professors in the United States revealed that psychologists were much more skeptical about psi than respondents in the humanities, the social sciences, or the physical sciences, including physics. They are more than twice as likely as respondents in other disciplines to assert that psi is impossible (34% to 16%) (Wagner & Monnet, 1979).

One frequently cited argument for being skeptical about psi is that there is no explanatory theory or proposed mechanism for psi phenomena that is compatible with current physical and biological

principles. Historically, of course, the discovery and scientific exploration of most phenomena have preceded explanatory theories, often by decades (e.g., the analgesic effect of aspirin; the anti-depressant effect of electroconvulsive therapy) or even centuries (e.g., electricity and magnetism, explored in ancient Greece as early as 600 BC, remained without theoretical explanation until the Nineteenth Century). The incompatibility of psi with our current conceptual model of physical reality may say less about psi than about the conceptual model of physical reality that most non-physicists, including psychologists, still take for granted—but which physicists no longer do.

As is widely known, the conceptual model of physical reality changed dramatically for physicists during the 20th Century, when quantum theory predicted and experiments confirmed the existence of several phenomena that are themselves incompatible with our everyday Newtonian conception of physical reality. Some psi researchers see sufficiently compelling parallels between certain quantum phenomena (e.g., quantum entanglement) and characteristics of psi to warrant considering them as potential mechanisms for psi phenomena. Moreover, specific mechanisms have been proposed that seek to explain psi effects with theories more testable and falsifiable than simple metaphor.

Although very few physicists are likely to be interested in pursuing explanations for psi, the American Association for the Advancement of Science (AAAS) has now sponsored two conferences of physicists and psi researchers specifically organized to discuss the extent to which precognition and retrocausation can be reconciled with current or modified versions of quantum theory (Sheehan, 2006, 2011).

Ironically, even if quantum-based theories of psi eventually do mature from metaphor to genuinely predictive models, they are still not likely to provide intuitively satisfying descriptive mechanisms for psi because quantum theory itself fails to provide such mechanisms for physical reality.

As physicist and Nobel Laureate Richard Feynman (1994) advised, “Do not keep saying to yourself... ‘but how can it be like that?’ because you will get...into a blind alley from which nobody has yet escaped. Nobody knows how it can be like that (p. 123).”

Meanwhile the data increasingly compel the conclusion that it really *is* like that.

Perhaps in the future, we will be able to make the same statement about psi.

References

- 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.
doi:10.1037/a0021524
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Feynman, R. (1994). *The character of physical law*. New York, NY: Modern Library.
- McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development*, *71*(1), 173-180.
- Mossbridge J, Tressoldi P and Utts J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. *Frontiers of Psychology* 3:390.
doi:10.3389/fpsyg.2012.00390.
- Richard, F. D., Bond, C. F. Jr. & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331-363.
- Sheehan, D. P. (Ed.) (2006). *Frontiers of time: Retrocausation—experiment and theory*. AIP Conference Proceedings (Vol. 1408), San Diego, California. Melville, New York: American Institute of Physics.
- Sheehan, D. P. (Ed.) (2011). *Quantum Retrocausation—theory and experiment*. AIP Conference Proceedings (Vol. 863), San Diego, California. Melville, New York: American Institute of Physics.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p*-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, *9*: 666-681. DOI: 10.1177/1745691614553988
- Steering Committee of the Physicians Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing physicians' health study. *The New England Journal of Medicine*, *318*, 262- 264.