On the plausibility of scientific hypotheses

D. Samuel Schwarzkopf
Experimental Psychology & Institute of Cognitive Neuroscience, University College London

Mossbridge and Radin review psychological and physiological experiments that purportedly show time-reversed effects. I discuss why these claims are not plausible. I conclude that scientists should generally consider the plausibility of the hypotheses they test.

Science seeks to explain a chaotic world by formulating lawful relationships that permit causal predictions. When I flip a coin, it will usually land on either heads or tails. The exact outcome depends on a multitude of factors and is difficult to predict – but I am confident that after many flips the number of heads and tails should be roughly equal. However, I watch as my friend Frank flips a coin and it keeps landing on heads. Something is obviously amiss.

It is possible that Frank is a wizard. In fact, if you knew Frank, you could be forgiven for thinking that. He looks and acts like a wizard. So, it is theoretically possible that Frank can magically force the coin to land on heads every single time. However, in spite of his arcane aura, this would not be my first hypothesis.

My first step would probably be to inspect the coin. Perhaps it shows heads on both sides? If that is not the case, I could give Frank another coin, one I know to be fair. If this one also lands on heads all the time, my next guess will be that Frank uses some kind of trick. Perhaps he is throwing the coin in a particular way that ensures it will land on heads? I would carefully watch how he flips the coin to see if I can spot anything unusual. I could compare his coin flipping movements to those of other people. I could enlist the help of modern technology and record his coin flips with a high-speed camera and play them back

in slow motion. When I get desperate to find an rational explanation, I might analyze magnetic fields and air pressure.

How long should I search for an explanation until I conclude that Frank is a wizard? Whatever the explanation for Frank's uncanny coin flipping ability, calling him a wizard is essentially admitting defeat. All we really know is that he can get coins to always land on heads. That is an interesting observation but it is not an explanation.

The hypothesis that Frank is a wizard is very *implausible*. I do not know any other wizards. I may have watched or read about some fictional wizards, and I know magicians who can perform elaborate magic tricks. But to the best of my knowledge, I have never witnessed the casting of any actual magical spells. I also have no idea how magic could physically work. I accept that I do not know everything about the universe – but I choose to go with what I *do* know. Therefore, my prior belief in the existence of magic, and in Frank being a wizard, remains extremely weak. Whatever his coin flipping abilities, it cannot convince me that he is a wizard. I want more conclusive evidence than that. For instance, if lots of wizards suddenly revealed they are capable of similar feats, the interpretation that Frank is one of them would seem far more likely.

It is the same with the scientific study of precognition. In the current issue of this journal, Mossbridge and Radin (henceforth, M&R) review studies testing the hypothesis that future events can influence the past. This includes experiments on precognitive dreams, lab experiments in which participants correctly guess events before they occurred, and so-called "presentiment" effects, physiological responses that manifest only *before* emotional stimuli. I will not address all the points they raise. I previously commented (Schwarzkopf, 2014) on their earlier meta-analysis on presentiment (Mossbridge, Tressoldi, & Utts, 2012). In my view, M&R fail to address my earlier concerns, in particular with regards to expectation bias or randomization procedures, but I do not want to dwell on those smaller points. Instead, I only want to discuss one fundamental issue: Are presentiment or precognition effects *plausible*?

I previously raised the concern that the time-reversed physiological responses to emotional stimuli reviewed by M&R are not *biologically* plausible. In conventional thinking, an emotional stimulus generates neuronal responses within the first few hundred milliseconds after it is shown to the participant. Slower physiological responses, such as pupil dilation, galvanic skin responses, and changes in neural blood flow then follow this neuronal response.

According to M&R, presentiment effects show similar differences in the response latency, but they are time-reversed: electrophysiological responses are reported to occur hundreds of milliseconds *prior* to the stimulus, while galvanic skin responses or hemodynamic changes occur even several seconds *earlier* (Mossbridge et al., 2012). Does this mean that all such events are mirrored back in time relative to stimulus onset (Bierman, 2010)? Does blood flow increase because several seconds *later* neurons in the brain will fire, and in turn they fire because even several hundred milliseconds *later* an emotional stimulus will appear? The main reason for hemodynamic responses in the brain is thought to be the metabolic demand caused by increased neuronal firing. Therefore, should these retro-causal electrophysiological responses not also *cause* hemodynamic consequences? If presentiment existed, the response to any stimulus would be a constant swamp of causal and retro-causal effects as well as their nonlinear interactions.

Similarly, it is implausible that participants can guess trials in a two-alternative forced choice task correctly at a rate of 51-53%, the rate Daryl Bem's precognition experiments reported (Bem, 2011). I am not a betting man but if this were true I would start a coin flip betting business. Even with such a miniscule winning margin, this would nevertheless soon turn a pretty healthy profit (Figure 1). Even if we accept ideas about quantum entanglement or other subatomic time-reversals as possible explanations, such effects should be *tiny*. Either Bem somehow amplified his participants' natural precognitive ability by several orders of magnitude, or his findings were the result of methodological flexibility and/or experimental artifacts instead. The latter is a far more plausible hypothesis.

In their review, M&R casually dismiss my earlier concerns with the plausibility of time-reversed phenomena like presentiment and precognition. According to them, the fact that

lab experiments found such effects directly demonstrates that they *are* plausible. This is a circular argument. A statistically significant observation does not prove that a hypothesis is true. The plausibility of a hypothesis depends on whether an observation is consistent with our current understanding of the world. I have little reason to believe that Frank is a wizard other than the fact that his coin flipping is unusual and that he kind of looks like one. I have no reason to believe that precognition is possible but for some anecdotes and a handful of parapsychology experiments with effect sizes that are really small – but at the same time far too large to be theoretically feasible.

M&R's argument is known as the base rate fallacy: No matter how strong the statistical evidence, if the hypothesis is impossible, it must necessarily be false. The p-value is irrelevant when the observed effect size cannot be observed under the alternative hypothesis. I cannot confidently claim that precognition or presentiment are *impossible*. I simply do not know enough about the universe to know this for certain. I am however extremely skeptical that such retro-causal effects exist. Critically, even if I accept that such effects are at least possible, the rate at which they can be observed in noisy psychology or physiology experiments must be nanoscopic, many orders of magnitude below those reported by these studies. The reported effects are not plausible under this hypothesis and thus alternative explanations are far more likely.

Therefore, I must disagree with M&R that we are dealing here with "scientific heresies of the first order." Rather this statement betrays a fundamental misunderstanding: there are no heresies in science. Dogma is antithetical to science and any assumption can be challenged. Critically, however, nobody should take you seriously without compelling evidence. Frank may very well be a wizard but unless you show me more conclusive evidence that wizards actually exist I remain doubtful. I am skeptical that precognition is even possible but I certainly will not be convinced of its existence by some implausible observations, no matter how significant the meta-analysis.

What evidence for precognition would I find compelling? The experimental test must be highly sensitive (much larger sample sizes and low-noise measurements) and provide rigorous control for methodological flexibility like p-hacking. In that regard, I applaud M&R's

call for pre-registered replications of these effects. Pre-registration provides a clear delineation of the confirmatory and exploratory aspects of a study. Statistical significance is only meaningful for the former. I would go one step further and suggest that such replications should be *Registered Reports* (https://cos.io/rr), a format enjoying increasing popularity in several psychology journals including the recently launched *Nature Human Behavior*. Here, the methods are refined in an initial stage of peer-review and data collection only commences when the methods have been finalized. However, even that does not control adequately for some of the problems that could skew the findings. To ensure that the results are convincing even to skeptics, the experiment should be conducted as an *adversarial collaboration* where skeptics and proponents of precognition effects work together to ensure the experiment is conducted in a way they both agree with. While such collaborations do not always end the disagreement between parties, both sides are given a chance to interpret the results – and the readers can make up their own mind about which hypothesis the evidence supports.

If all these steps have been taken and precognition findings nonetheless replicate in a set of homogenous replications, I will accept that there is a result worthy of an explanation. However, even such a finding still does not mean that precognition exists. If the effect size is similar to Bem's reports of 51-53% correct, then it is simply not *plausible* that this occurs in the general population and everyday situations. At most, this would imply that precognition can only be demonstrated in these particular experimental contexts, which seems rather unlikely. The onus then is on proponents of the precognition hypothesis to show experimentally what makes this effect so unstable. If they cannot do so, methodological artifacts or other uncontrolled flexibility remain a more plausible alternative explanation.

In general, the burden of proof must always lie with the one making a claim. Therefore, it falls on proponents of a novel hypothesis to provide compelling evidence for it. Moreover, a fundamental principle of scientific research is that a hypothesis should be *falsifiable*. Before setting out to test a new hypothesis, an investigator should always ask themselves what evidence could *disprove* this hypothesis. If they cannot answer this question, the hypothesis is probably not worth testing. To my knowledge, proponents of precognition have yet to provide an answer to this question.

But let me be clear: my problem with the research on precognition is not with its fringe nature. Instead it is with the approach and the interpretation of these findings. This is not a problem limited to parapsychology but it plagues a lot of scientific research. The precognition effects reported in these studies are not plausible but neither are claims that unscrambling words related to old age can make participants walk down a corridor a second more slowly than controls (Bargh, Chen, & Burrows, 1996). Given the messy nature of human behavior it seems very unlikely that a simple psychology experiment can have such a profound effect. It should therefore not surprise anyone when such findings fail to replicate (Doyen, Klein, Pichon, & Cleeremans, 2012). The same principle must apply to reports of gravitational waves, discoveries of arsenic microbes, brain-behavior correlations, and even simple psychophysical tests of visual perception.

We can all do a lot better. We should put our hypotheses to much greater scrutiny. If you observe an effect, you must ask whether it is plausible under the hypothesis you are testing. Extraordinary claims require extraordinary evidence. And always ask yourself, what would convince you that you are wrong. Mossbridge and Radin clearly challenge our current science – just not in the way they seem to think.

References

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244.

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. https://doi.org/10.1037/a0021524

Bierman, D. (2010). Consciousness-induced restoration of time symmetry (CIRTS). A psychophysical theoretical perspective. Journal of Parapsychology, 24, 273–300. *Journal of Parapsychology*, *24*, 273–300.

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PloS One*, *7*(1), e29081. https://doi.org/10.1371/journal.pone.0029081

Mossbridge, J., Tressoldi, P., & Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. *Frontiers in Psychology*, *3*, 390. https://doi.org/10.3389/fpsyg.2012.00390

Schwarzkopf, D. S. (2014). We should have seen this coming. *Frontiers in Human Neuroscience*, *8*, 332. https://doi.org/10.3389/fnhum.2014.00332

Figure captions

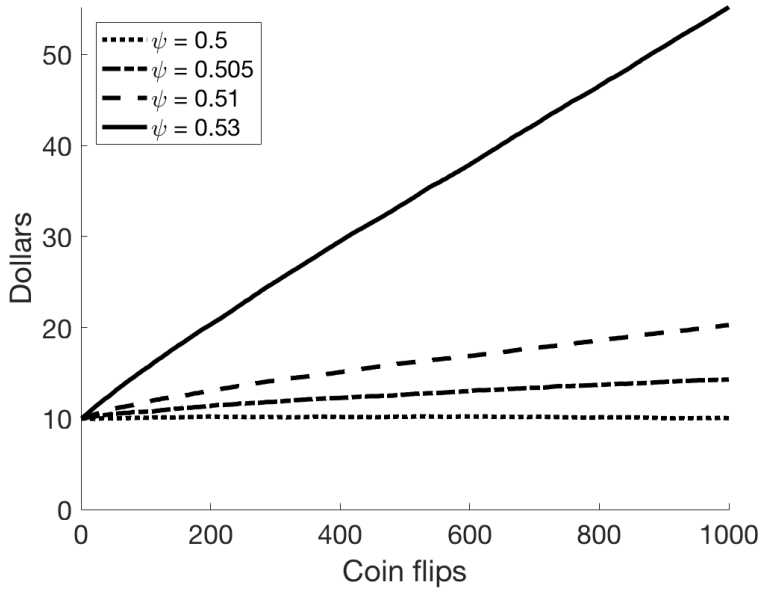

Figure 1. Simulated universes in which typical reported precognition effects exist. I start with $10. For every coin flip, I bet $1 that I can guess the outcome. If I guess correctly, I get

back $2 and thus win $1. If I fail, I lose my bet. I keep flipping the coin 1000 times or until I run out of money. The curves show the amount of money I have, plotted against the number of coin flips (averaged across 10,000 simulations). Different line styles denote different "natural precognition rates". Without any precognition (ψ=0.5), I would not win any money. However, even with tiny precognition effects ($0.5<\psi\leq0.53$) I would turn a profit. See https://doi.org/10.6084/m9.figshare.4879835.v1 for the Matlab code.