

## Kahneman's Comment

Daniel Kahneman February 14, 2017 at 8:37 pm

From Daniel Kahneman

I accept the basic conclusions of this blog. To be clear, I do so (1) without expressing an opinion about the statistical techniques it employed and (2) without stating an opinion about the validity and replicability of the individual studies I cited.

What the blog gets absolutely right is that I placed too much faith in underpowered studies. As pointed out in the blog, and earlier by Andrew Gelman, there is a special irony in my mistake because the first paper that Amos Tversky and I published was about the belief in the “law of small numbers,” which allows researchers to trust the results of underpowered studies with unreasonably small samples. We also cited Overall (1969) for showing “that the prevalence of studies deficient in statistical power is not only wasteful but actually pernicious: it results in a large proportion of invalid rejections of the null hypothesis among published results.” Our article was written in 1969 and published in 1971, but I failed to internalize its message.

My position when I wrote “Thinking, Fast and Slow” was that if a large body of evidence published in reputable journals supports an initially implausible conclusion, then scientific norms require us to believe that conclusion. Implausibility is not sufficient to justify disbelief, and belief in well-supported scientific conclusions is not optional. This position still seems reasonable to me – it is why I think people should believe in climate change. But the argument only holds when all relevant results are published.

I knew, of course, that the results of priming studies were based on small samples, that the effect sizes were perhaps implausibly large, and that no single study was conclusive on its own. What impressed me was the unanimity and coherence of the results reported by many laboratories. I concluded that priming effects are easy for skilled experimenters to induce, and that they are robust. However, I now understand that my reasoning was flawed and that I should have known better. Unanimity of underpowered studies provides compelling evidence for the existence of a severe file-drawer problem (and/or p-hacking). The argument is inescapable: Studies that are underpowered for the detection of plausible effects must occasionally return non-significant results even when the research hypothesis is true – the absence of these results is evidence that something is amiss in the published record. Furthermore, the existence of a substantial file-drawer effect undermines the two main tools that psychologists use to accumulate evidence for a broad hypotheses: meta-analysis and conceptual replication. Clearly, the experimental evidence for the ideas I presented in that chapter was significantly weaker than I believed when I wrote it. This was simply an error: I knew all I needed to know to moderate my enthusiasm for the surprising and elegant findings that I cited, but I did not think it through. When questions were later raised about the robustness of priming results I hoped that the authors of this research would rally to bolster their case by stronger evidence, but this did not happen.

I still believe that actions can be primed, sometimes even by stimuli of which the person is unaware. There is adequate evidence for all the building blocks: semantic

priming, significant processing of stimuli that are not consciously perceived, and ideomotor activation. I see no reason to draw a sharp line between the priming of thoughts and the priming of actions. A case can therefore be made for priming on this indirect evidence. But I have changed my views about the size of behavioral priming effects – they cannot be as large and as robust as my chapter suggested.

I am still attached to every study that I cited, and have not unbelieved them, to use Daniel Gilbert's phrase. I would be happy to see each of them replicated in a large sample. The lesson I have learned, however, is that authors who review a field should be wary of using memorable results of underpowered studies as evidence for their claims.

## **Reply**

Dr. R February 14, 2017 at 8:57 pm

Dear Daniel Kahneman,

Thank you for your response to my blog.

Science relies on trust and we all knew that non-significant results were not published, but we had no idea how weak the published results were.

Nobody expected a train-wreck of this magnitude.

Hindsight (like my bias analysis of old studies) is 20/20. The real challenge is how the field and individuals respond to the evidence of a major crisis.

I hope more senior psychologists will follow your example and work towards improving our science.

Although we have fewer answers today than we thought we had five years ago, we still have many important questions that deserve a scientific answer.