Fixing the Replication Crisis: The Need to Understand Human Psychology

November 25, 2019

In the past 15 years, there has been enormous progress in documenting problems with the credibility of research findings, not just in our own field but also in many areas of science. Metascience studies have helped us quantify the extent of the problem and have begun to shed light on the underlying causes. We need to move now to a focus on fixing the problems rather than just illustrating them. But can this be done?

Many of the problems currently under discussion have been known for decades. For instance, in 1976, Michael Mahoney wrote a book called *Scientist as Subject*: *The Psychological Imperative*, in which he discussed the bias that reviewers show toward their favored ideas. He gave evidence from an experiment in which he asked 75 reviewers to referee journal articles that were identical except for the results, which could be positive, null, or mixed. He found that reviewers were biased against results that did not support their theoretical position. As a consequence, he proposed a new approach to journal reviewing that anticipated by some 30 years the idea of a Registered Report, arguing that "manuscripts should be evaluated solely on the basis of their relevance and their methodology. Given that they ask an important question in an experimentally meaningful way, and they should be published — regardless of their results." (Mahoney, p. 105) Yet his demonstration of bias and his suggested solution were overlooked, and we continue to see strong evidence of publication bias in our journals. Many editors are reluctant to accept null results, regardless of how well-designed a study is.

In a similar vein, Jacob Cohen's (1969) exhortation to do adequately powered studies remains largely ignored.

Why do these problems with scientific practice persist, and why aren't we doing more to solve them?

Most responses to this question focus on either training or incentives. Those who advocate training posit that people do bad science because they confuse it with good science. If we train them better, they will improve. Others point out that people are led astray by skewed incentives regardless of training. Rewarding scientists for publishing in high-impact journals and acquiring large amounts of grant income will lead people to chase these proxy indicators of good science in a way that can corrupt the scientific process.

I agree that training and incentives are important issues to tackle if we want to have a hope of improving science, but I think we need also to take into account a third factor: human cognitive biases. Misunderstanding of statistics, and the incentive structure that has evolved, have their roots in human cognition. As I discuss in a recent article (Bishop, 2019), scientific thinking is not natural for humans: To be good scientists, we often have to actively inhibit our normal ways of thinking. Amos Tversky and Daniel Kahneman's (1971) article on "belief in the law of small numbers" illustrated how bad we are at appreciating the impact of sampling error on estimates based on small numbers. This, I think, explains

why we can keep explaining power analysis to researchers, and why they will continue to not take it seriously. *p*-hacking reflects a different aspect of statistical misunderstanding: On the one hand, there is a failure to appreciate that a *p* value cannot be interpreted out of context (de Groot, 2014), but I would argue that there is the added tendency to regard errors of omission as less serious than errors of commission (Haidt & Baron, 1996).

Thus, failing to report null results, even though they are an essential part of the context of interpretation of a *p* value, is regarded as far less serious than tweaking a *p* value to push it into significance. I term this "moral asymmetry," and I propose that it also plays a role in publication bias (where failing to report null findings is seen as innocuous) and the equally serious though less-documented tendency for citation bias (i.e., writing reviews that simply omit evidence that does not fit).

A final cognitive bias relates to our need for narrative to structure events. This was noted by Bartlett (1932) in his writings on "reconstructive remembering": our tendency to filter information in perception and memory to fit our existing schemata. Bartlett emphasised the beneficial consequences: We avoid information overload and can focus on what is meaningful. Science would not advance at all if we just had mountains of unstructured data: We need to make sense of observations and use our theoretical understanding to guide our interpretation. But this reconstructive tendency has a negative side. It leads us to ignore facts that don't fit and to present our research as if it told a much neater story than is usually the case.

Overall, I suggest that no amount of training in statistics or exhortations to behave differently will be effective in tackling the replication crisis unless we understand the cognitive basis of the biases that lead us astray. Fortunately, as psychological scientists, we are well-placed to do this, given our rich history of research on human cognition.

Look for more commentaries about metascience from leading researchers in upcoming issues of the Observer and follow the series online.

References

Barber, T. X. (1976). *Pitfalls in Human Research*. New York, NY: Pergamon Press.

Bartlett, F. C. (1932). *Remembering. A study in experimental and social psychology*. New York, NY: Cambridge University Press.

Bishop, D. V. M. (2019). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 45th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*. Advance online publication. http://doi.org/10.1177/1747021819886519

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

De Groot, A. D. (2014). The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas]. *Acta*

Psychologica, 148, 188–194. http://doi.org/10.1016/j.actpsy.2014.02.001

Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology, 26*, 201–218.

Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative.* Cambridge, MA: Ballinger.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.
http://doi.org/10.1037/h0031322