

Robust Science Depends on Understanding the Science of Humans

December 30, 2016



My colleague Howard C. Nusbaum is on leave from the University of Chicago Department of Psychology, serving as the Director for the Division of Behavioral and Cognitive Sciences in the National Science Foundation (NSF) Directorate for Social, Behavioral, and Economic Sciences. I invited him to weigh in on the goals of reliability, validity, and replicability from his vantage point at the NSF, and to discuss the special role that psychological science can play not only in achieving these goals, but also in understanding why they are so difficult to achieve.

-APS President Susan Goldin-Meadow

Science is a method of generating knowledge and testing beliefs; it trumps authority by empirical evaluation and depends on reliability and validity to uphold that knowledge. But science also is a human enterprise: Whether in psychology or neuroscience, physics or chemistry, studies are designed, conducted, and reported by people. Even with computer-controlled experiments, humans bear responsibility for the findings. When humans are involved, errors will occur. Some errors result from cognitive biases in decisions and judgment (Gilovich, Griffin, & Kahneman, 2002; Tversky & Kahneman, 1974), including confirmation bias (e.g., Nickerson, 1998); others occur by accident,

oversight, or carelessness; still others may be motivated.

A “motivated” error occurs when results are at odds with reality and are produced with the intention of distorting or fabricating the analysis for reasons independent of objective evidence, whether because of conviction or gain (Broad & Wade, 1982). Diagnosing motivated error is difficult. Allegra Goodman’s novel *Intuition* (2006) illustrates how personal and professional motivations can muddy the waters of scientific knowledge when error occurs. Selective data reporting is a turning point in the novel. In much the same way, William Broad and Nicholas Wade (1982) discussed how Robert A. Millikan received the Nobel Prize for demonstrating quantal electrical charge, which he did by selective reporting. In the end, though, replication wins out as the natural scientific corrective process.

Psychological science has always been especially mindful of the tools of reliability and validity as a consequence of our intellectual history. Psychology moved from analytic introspection to intersubjective testability to develop a science relying on objective and systematic methodology. This methodology puts psychological science on the same objective footing as research in physical sciences. Acceptance of this regimen is why we object to the false distinction of putatively hard (e.g., physics) and soft (e.g., psychology) sciences. The scientific method establishes parity, and the target of understanding phenomena that are not directly observable — whether states of mind or dark matter — certainly does not cleave the sciences apart.

Although we’re well aware of the controversies over replication in psychological science, it is important to remember that all sciences suffer the same issues. Physics has dealt with controversies over cold fusion and faster-than-light particles, but ultimately scientific theory and replication led to clarity. However, replication is not always the answer. Consider Prosper-René Blondlot’s 1903 discovery of N-rays (Broad & Wade, 1982). This discovery, a physical phenomenon, was replicated by a physical process in hundreds of papers (Simon, 2014; Tretkoff, 2007), but in spite of replications, there were skeptics. A skeptical physicist visited Blondlot and by simple intervention showed that the only real phenomenon was observation bias.

When theory and knowledge impel replication, they can redress scientific error, but reliability and validity are not the same. Statistical analyses raise the question of whether some results are too good to be true (Francis, Tanzman, & Matthews, 2014). This kind of analysis essentially questions whether reported results are statistically plausible. The results do not indicate error as such but simply flag what appear to be improbable findings that might not be replicated.

This is not a problem for psychology alone and has arisen in genetics (Francis, 2014). Gregor Mendel’s data were also too good to be true (Broad & Wade, 1982; Gelman, 2012). But Mendelian genetics withstand the test of time. Statistics bolster an argument but do not represent the whole truth of a result. Given that there are many ways for errors to distort research and derail progress, we need to understand how social, cultural, and psychological forces work in science.

Science must generate believable and robust knowledge. The NSF Advisory Committee to the Social, Behavioral, & Economic Sciences Directorate (SBE) established a Subcommittee on Replicability in Science. Their report (Bollen et al., 2015) defined robust findings as reproducible, replicable, and generalizable with clear definitions. This report was a call to support new robust and reliable science,

and SBE has posted a “Dear Colleague Letter” (DCL; Cook, 2016) announcing support for research on failures of robustness, methods to improve robustness, training to enhance robustness of research, and support for replications/generalizations of key SBE studies. Given that replicability concerns extend to other social sciences (Camerer et al., 2016), SBE is committed to improving robustness of SBE sciences. A DCL from the Directorate for Computer and Information Science and Engineering (Kurose, 2016) announces support for reproducibility in computing and communications research.

Science depends on credibility. There are, however, many ways that findings can fail replication, and not all compromise validity. Science requires humility given uncertainty and unknowns outside current research. The SBE DCL supports increasing robust research and greater reflection about robust science generally and hopes to lead to a wiser approach to research. The problem of robust science is not unique to the social and behavioral sciences; it inheres in all sciences and does so because all science is conducted by scientists — physicists and geneticists are human and thus subject to the social and psychological forces that can lead research astray. That alone creates a unique and important responsibility for understanding the conduct of robust science in the social and behavioral sciences. æ

References

- Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., Olds, J. L., & Dean, H. (2015). *Report of the Subcommittee on the Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, & Economic Sciences*. Retrieved from https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf
- Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. New York, NY: Simon & Schuster.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., ... Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351, 1433–1436.
- Cook, F. L. (2016). Dear Colleague Letter: Robust and reliable research in the social, behavioral, and economic sciences. Retrieved from <https://www.nsf.gov/pubs/2016/nsf16137/nsf16137.jsp>
- Francis, G. (2014). Too much success for recent groundbreaking epigenetic experiments. *Genetics*, 198, 449–451.
- Francis, G., Tanzman, J., & Matthews, W. J. (2014). Excess success for psychology articles in the journal *Science*. *PLOS One*, 9, e114255.
- Gelman, A. (2012). Gregor Mendel’s suspicious data [Blog post]. Retrieved from <http://andrewgelman.com/2012/08/08/gregor-mendels-suspicious-data/>
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- Goodman, A. (2006). *Intuition*. New York, NY: The Dial Press.

Kurose, J. (2016). Dear Colleague Letter: Encouraging reproducibility in computing and communications research. Retrieved from <https://www.nsf.gov/pubs/2017/nsf17022/nsf17022.jsp>

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.

Simon, M. (2014). Fantastically wrong: The imaginary radiation that shocked science and ruined its ‘discoverer.’ *Wired*. Retrieved from <https://www.wired.com/2014/09/fantastically-wrong-n-rays/>

Tretkoff, E. (2007). This month in physics history: September 1904: Robert Wood debunks N-rays. *American Physical Society*. Retrieved from <https://www.aps.org/publications/apsnews/200708/history.cfm>

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.