Expert Opinion

Editor: Christina M. Finello, JD, Ph.D.

The Philosophy of Science on Testing Research and Clinical Questions

Tess M.S. Neal, Ph.D.
National Science Foundation Interdisciplinary Postdoctoral Research Fellow
University of Nebraska Public Policy Center
Incoming Assistant Professor, Interdisciplinary College of Arts & Sciences
Arizona State University

The focus of this column is on applying the philosophy of science to both experimental and clinical [including forensic] psychological work. Specifically, it is about how to develop and test a hypothesis or clinical intuition in a logically defensible way - either for a research study or for a clinical/forensic assessment. Before reading further, please complete this exercise:

There is a set of cards. On Side 1 is a colored shape and on Side 2 is a letter. The following hypothesis is proposed: *"Whenever there is a ▲ on Side 1, there is a Z on Side 2."* Please look at these 4 cards, two with Side 1 exposed and two with Side 2 exposed:

The question is: *What is the minimum number of cards, and which ones, should you turn over to test the hypothesis?*

Most people (even trained scientists and professionals) have trouble correctly answering this question. The answer is 2 cards - the "blue triangle" and "Y" cards. Finding a Y on the back of the blue triangle would allow you to reject the hypothesis by falsifying it, as would finding a blue triangle on the back of the "Y" card. No other possibilities would reject the hypothesis (task adapted from Wason, 1968; see also Neal & Grisso, 2014). Turning over the green square is not helpful because it is irrelevant to the hypothesis no matter what is on the other side. Neither will turning over "Z" allow you to reject the hypothesis; seeing a blue triangle on the reverse side would vacuously confirm the hypothesis, and seeing a green square would not tell you anything about the hypothesis. If you thought the "Z" should be turned over (most people do), you made a common error that demonstrates just the point of this column.

Turning over the Z card is consistent with a cognitive error called the *positive test strategy* (Kayman & Ha, 1987). This mental heuristic leads to testing hypotheses by searching for evidence that has the best chance of *verifying* the hypothesis, rather than those that have the best chance of *falsifying* it. This bias

is pervasive, as it is easy and relatively automatic for people to think in terms of verification, or searching for supporting data (MacCoun, 1998; Nickerson, 1998).

This confirmation/verification method of empiricism has had a strong presence in the history of science and influenced the development of the scientific method. It was not recognized as a problem with the scientific method until relatively recently. For instance, verification is evident in Francis Bacon's writings about induction in the seventeenth century (1620, *Novum Organum*) and Auguste Comte's positivism in the nineteenth century (1848, *A General View of Positivism*). It wasn't until the late 1950s and early 1960s that postpositivist philosophers of science such as Karl Popper (1959) and Thomas Kuhn (1962) recognized and explained the problems of verification to revise the modern scientific method.

The revised scientific method requires that scientists and scientist-practioners develop falsifiable hypotheses (i.e., able to be disproven with a single observation) and then explicitly attempt to disprove the hypotheses by searching for disconfirming, rather than confirming, evidence. This is hard for people – even well-trained trained scientists and clinicians – to do, but is nevertheless the logical and most defensible way to test hypotheses because it is a safeguard against confirmation bias.

Consider an illustrative example. Let's say our hypothesis is that all swans are white. We go out and collect data, observing a sample of 100 different swans and coding their color. Now let's say our hypothetical results show 100 white swans. In this case, we would say our hypothesis was confirmed. We might then think we discovered some "truth" about the universe – that all swans are indeed white. Even if we saw a *million* white swans, it wouldn't change our interpretation (though it might make us pretty confident). However, it would only take a *single observation* of a black swan to disprove our hypothesis. If we saw a black swan, we would know immediately that what we thought was a truth in the world ("all swans are white") was actually wrong.

This example demonstrates that the absence of disconfirming evidence (seeing no non-white swans in a sample of 100, or even in a million swans) cannot be equated with proof. No number of confirming observations can "prove" a hypothesis or theory true. And it takes just one observation to prove a (falsifiable) theory false. Albert Einstein was ahead of his time by realizing that "no amount of experimentation can prove me right; a single experiment can prove me wrong" (Rao, 2001, p. 2244). The "fittest" theories are those that survive repeated disconfirmation attempts – those theories may better approximate "truth" than theories that have simply been confirmed.

Researchers and clinicians should explicitly try to emphasize searching for disconfirming rather than confirming evidence in designing tests of their hypotheses. Researchers formulate a priori hypotheses about what they expect to find prior to collecting and analyzing data. And clinicians form clinical hypotheses about what they think might be going on in a particular case, subsequently collecting more information to test those clinical hypotheses as they develop their diagnoses and clinical opinions in the case.

Researchers should design their studies by explicitly crafting falsifiable hypotheses with methods to disprove them. For instance, a researcher might be interested in how juvenile age affects judicial decision making. The hypothesis might be that the younger the juvenile, the lower judicial perceptions of culpability and dangerousness would be, based on an "innocence of youth" type of heuristic. To effectively test this falsifiable hypothesis, the researcher would need to compare judicial ratings of culpability and dangerousness for younger and older juvenile defendants, searching for evidence that judges might actually perceive youth who start offending younger as *more* culpable and dangerous than older youth.

A clinician referred a forensic case in which the 45-year-old defendant was charged with

torturing and killing animals might hypothesize the defendant meets the diagnostic criteria for Antisocial Personality Disorder. This clinician might look for evidence of persistent failure to obey laws and social norms such as numerous arrests for criminal behavior, a pattern of irresponsible behavior, a pervasive pattern of lying to and deceiving others, repeated fights, and impulsivity. While finding information about each of these symptoms incrementally confirms the hypothesis, it is by looking for information about the defendant's behavior before 15 years of age that could swiftly disprove the hypothesis. If these symptoms only began to emerge during middle adulthood, it would rule out the diagnosis of Antisocial Personality Disorder – even if supporting information about each and every one of those other behaviors was found. If this occurred, the clinician would then need to revise the hypothesis; perhaps the defendant sustained a traumatic brain injury at 40 years of age that changed his personality and could better explain his pattern of symptoms. With the revised hypothesis, the clinician would then set out to try to disprove it – continuing on until the most robust clinical hypothesis could not be disproven.

In sum, appropriate methods for testing research and clinical hypotheses rely on a conception of knowledge as a particular *approach* to generating it. Described wonderfully by Richard Feynman (1985), a Nobel-prize willing theoretical physicist, science is "bending over backward to prove ourselves wrong." Researchers and clinicians should approach their work in this spirit.

Author Bio:

Tess M.S. Neal, Ph.D., is a National Science Foundation interdisciplinary postdoctoral research fellow at the University of Nebraska Public Policy Center. She will be joining the Interdisciplinary College of Arts and Sciences at Arizona State University as an assistant professor of psychology in fall 2015. She obtained her Ph.D. in clinical psychology at the University of Alabama and completed a clinical-forensic postdoctoral residency at the University of Massachusetts Medical School. Her research interests focus on basic human judgment and decision making in applied contexts. http://ppc.unl.edu/staff/tess-m-s-neal/

References

Feynman, R.P. (1985). *Surely you're joking, Mr. Feynman: Adventures of a curious character.* New York: Norton.

Klayman, J. & Ha, Y.W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228. doi: 10.1037/0033-295X.94.2.211

Kuhn, T.S. (1962). *The structure of scientific revolutions.* Chicago, IL: University of Chicago Press.

MacCoun, R.J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology, 49*, 259-287. doi: 10.1146/annurev.psych.49.1.259

Neal, T.M.S. & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law, 20,* 200-211. doi:10.1037/a0035824

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220. doi: 10.1037/1089-2680.2.2.175

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Rao, C.R. (2001). Statistics: Reflections on the past visions of the future. *Communication in Statistics: Theory & Methods, 30*, 2235-2257. doi: 10.1081/STA-100107683