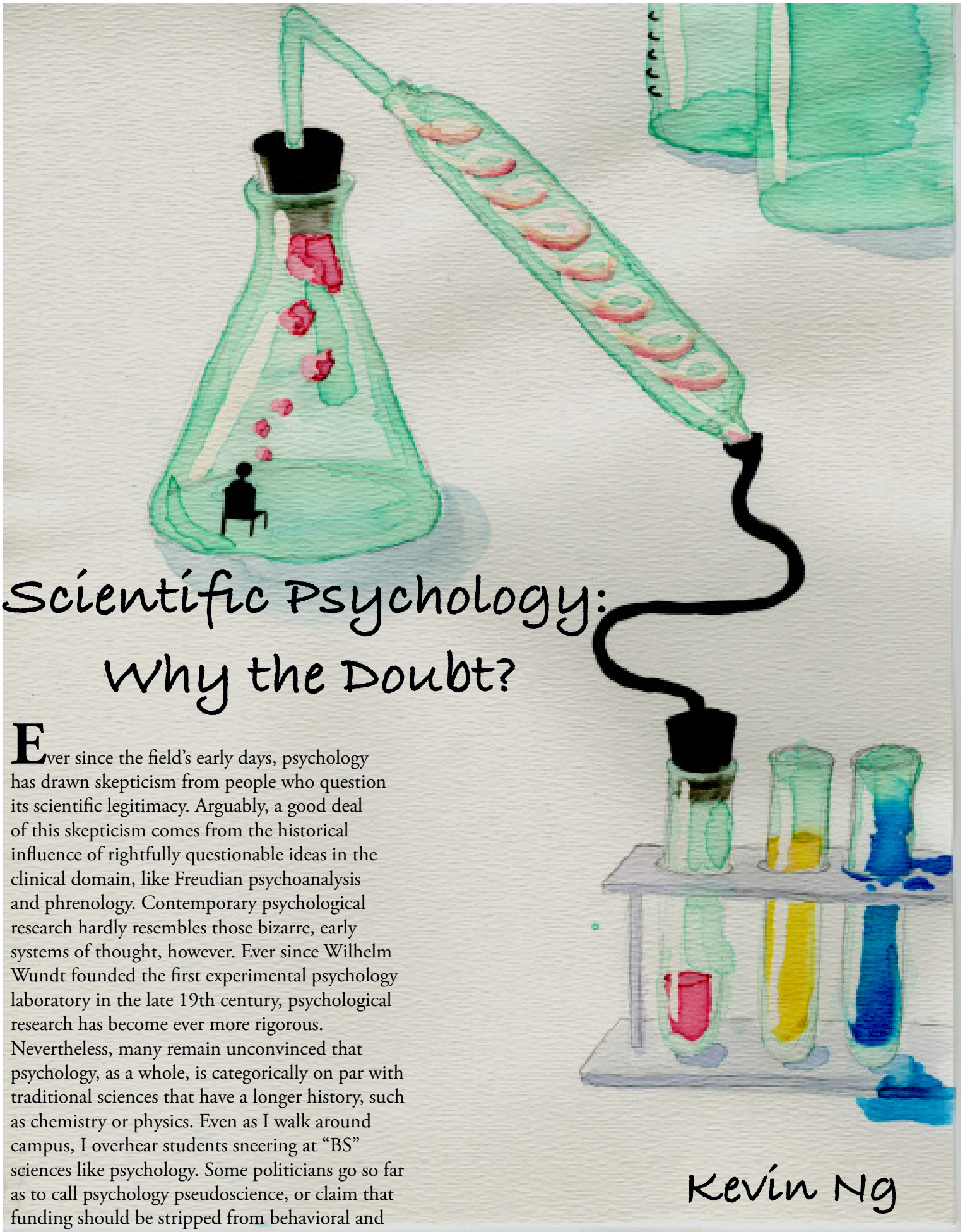


Scientific Psychology: Why the Doubt?

Kevin Ng

Ever since the field's early days, psychology has drawn skepticism from people who question its scientific legitimacy. Arguably, a good deal of this skepticism comes from the historical influence of rightfully questionable ideas in the clinical domain, like Freudian psychoanalysis and phrenology. Contemporary psychological research hardly resembles those bizarre, early systems of thought, however. Ever since Wilhelm Wundt founded the first experimental psychology laboratory in the late 19th century, psychological research has become ever more rigorous. Nevertheless, many remain unconvinced that psychology, as a whole, is categorically on par with traditional sciences that have a longer history, such as chemistry or physics. Even as I walk around campus, I overhear students sneering at "BS" sciences like psychology. Some politicians go so far as to call psychology pseudoscience, or claim that funding should be stripped from behavioral and

social science research. For example, Senator Tom Coburn of Oklahoma had already managed to block NSF funding for political science research in 2013. If the definition of science could so drastically affect a community's eligibility for research funds, then the question arises: what is science?

Philosophers have debated over the definition of science for millennium, and they have yet to agree upon a strict, concise definition. Solving the demarcation problem, as it's called, is a challenge that remains open to future philosophers. However, most of us practicing scientists do not have time to delve into a deep study of Philosophy of Science – Popper's falsificationism, Kuhn's Paradigm Shifts, or Post-positivist Theory. These strange phrases refer to philosophical frameworks which try to pinpoint what is essential to the scientific process. Falsificationism was developed by a Viennese philosopher named Karl Popper, who states that only ideas that can be shown to be wrong are to be considered scientifically meaningful (Popper went a bit further and said that, even if an idea has survived multiple attempts at falsification, our confidence in the idea should not be increased; we should just say it has not yet been falsified and no more than that). Another philosopher, Thomas Kuhn, emphasized the social aspects of the scientific enterprise; he said that scientific progress requires that scientists not question the fundamental ideas, or paradigm, of their field so that the field can have the chance to mature. However, when enough observed anomalies, or things that don not fit their guiding framework, accumulate, the basic ideas are back out onto the table for discussion in what is called a Paradigm Shift. After a new framework emerges from the chaos, science proceeds as usual. Post-positivism refers to a broad set of sociological ideas that critique the supposition that how scientists think and how common sense works are not different; Post-positivists tend to believe all observation is fallible and cannot be value-free, that scientific ideas are not independent of the social, economic, and political forces that surround it, and that science is actually the endeavor to reach the unattainable goal of acquiring truth without subjectivity.

Of course, these are gross generalizations of rather complex philosophies. Yet, to reiterate, few scientists have the time to dig deeper into these ideas on the basis of busy schedules – some may not even have the interest. It is therefore helpful to have a "rule of thumb" definition of science as something to reference in conversations, so that we don't stare with a blank face at the task of defining science. There are enough similarities across disciplines to allow at least some formulation of a working set of essential characteristics. Keith Stanovich, Professor of Applied Psychology at the University of Toronto and author of How to Think Straight About Psychology, offers such a definition. In his primer on critical thinking in psychology, he lays out three broad principles that define scientific inquiry.

1. Systematic empiricism
2. Public verifiability
3. Exploration of falsifiable hypotheses

The first is systematic empiricism. Many believe "empirical" means experimental, but it can actually mean something a bit broader than that. "Empirical" comes from the Latin transliteration of a Greek inflection of "empeiria", meaning "experience." This makes sense, since scientists rely on observation; we observe a phenomenon by experiencing it through one or more of our five senses. However, relying only on raw, unstructured experiences is not a sufficient condition, otherwise we would all be scientists all the time. Furthermore, writing down every observation you make from dawn till dusk will result in a huge list of facts at the end of the day, but it will not yield a more fundamental or generalizable understanding of the world. Scientists structure their observations so as to uncover something underlying about a phenomenon.

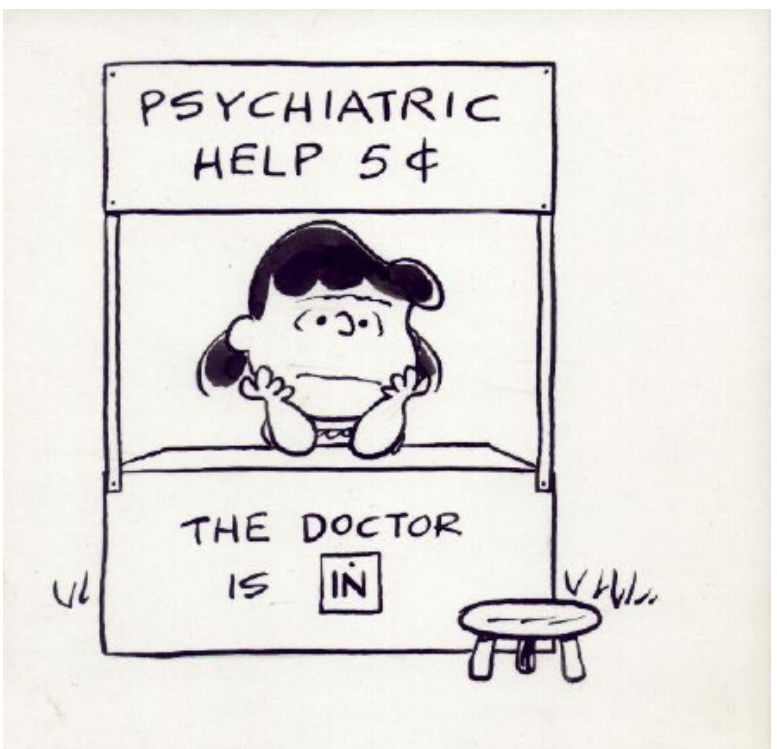
The second criterion is public verifiability. Ideas trapped inside the head of someone who drew conclusions from a study have little utility for the rest of the scientific community. They must be submitted to the scientific community for scrutiny, discussion, and criticism. Two mechanisms act as quality control for submitted information: peer review and replication. Peer review standards vary from journal to journal, and the open-access movement is adding a new dimension to it. Nonetheless, what is important is that information that has been vetted by peer review has met a minimal standard of scrutiny, even if it is not necessarily correct. Without some level of scrutiny by the relevant experts in the field, journals run the risk of blatant misinformation or baseless assertions being placed side-by-side corroborated claims. Peer review is not meant to be the final arbiter of what is true and what is not; it is simply intended to make sure that well-supported ideas propagate through the field. Replicating studies is what enables scientists to sort out the true findings from the spurious ones. It gives other researchers a way to make sure a certain finding was not the result of the biases or errors of a particular set of scientists.

Continued from page 15

The third criterion is the exploration of falsifiable hypotheses. Scientists address ideas that can produce confirming or damning evidence. If a theory poses certain hypotheses, and those hypotheses consistently fail to be supported, then the theory should be either adjusted or discarded. Either way, it has been falsified. This is a good thing; consistent failures to fulfill predictions mean that certain ideas are being supported while others are not. While we can't be absolutely certain in this process of elimination, it does free us from having to be equally open-minded to all empirical claims. Scientific theories need to be presented such that they can be shown to be wrong. However, the boundary between the falsifiable and the unfalsifiable can be blurry; as empirical methods improve, questions can drift into the scientific domain. For example, many years ago, historians believed that there was no way to figure out whether or not Thomas Jefferson was the father of a mixed-race slave boy. As the field of Genetics became more technologically advanced, however, the question gradually became recognized as testable. By the end of the 20th century, knowledge about DNA and genetic technology advanced to the point where methods for verifying bloodlines were available, and Jefferson's fatherhood status had confirming evidence at last.

Given Stanovich's "rule of thumb" definition of science, how does psychology fare? There is definitely systematic empiricism, ranging from case reports to controlled trials. Psychologists also conduct correlational studies, which are less rigorous than experimental studies, but they are still a form of data collected from structured observations. Astronomy is an example of a science where controlled experiments can be rare – it is rather impossible to experimentally manipulate asteroids or supernovae. There are also a great number of peer-reviewed psychology journals, a reflection of the diversity of topics the field covers.

An interesting issue comes up with regard to falsifiability and replication. In November of 2012, the journal Perspectives on Psychological Science published a series of articles addressing a "crisis of replication" in psychology. Many were concerned that replication studies did not occur often enough in many of psychology's sub-fields. This is problematic, because replication is one of the classic mechanisms of science by which false findings are filtered out from the true findings. In one review, researchers tried approximating the number of replication studies in the field by searching the publication history of 100 high-impact psychology journals for the stem "replicat-". Only 1.6% of journal articles in the search had the stem; an even smaller percentage of that 1.6% were verified as actual replications. The authors note that psychology is not alone in this regard, as other studies found similar replication rates in other fields, like business and certain medical subfields. There are many likely contributors to why replication studies may not occur in a field, including lack of funding, differences in what research is prioritized, differences in academic incentives, etc. A change in any one of these contributors might lead to a change in how psychology is practiced. That is, if funding agencies recognized the value of psychological research, they might provide more funding, which would lead to more studies overall and perhaps a higher percentage of replication studies. If the academic culture shifted to incentivize repeating other researchers' work, the practice of psychological research might entail direct replication studies more regularly. A change in the cultural norms of the practice of conducting research may be what it takes to get more replication happening in these fields.



Another issue that is hardly restricted to psychology is the problem of publication bias. In 1959, a statistician named Theodore Sterling was looking at fields that commonly used null hypothesis statistical significance testing and found that around 97% of a sample of four high-impact psychology journals reported positive, or statistically significant, findings. He explained that this was suspiciously large; even for actual phenomena, one expects at least some null findings, just by chance. Human

behavior is probabilistic, so where were all the studies in which no relationship between variables was found? He did the same thing more than three decades later and still found a huge proportion of positive findings. Just five years ago, Daniele Fanelli did the same thing; he found psychological findings to be much more likely to be positive than findings in the natural sciences, though the social sciences in general seemed to exhibit a similar pattern.

In an academic culture where researchers feel the pressure to publish positive findings, and where no one is going to get tenure any time soon by reporting that nothing was found or repeating someone else's work, these results are hardly surprising. Yet the danger is eminent. A scientist can run a study once and find nothing, but if they run the same study 19 more times, then by chance one could give them a publishable, significant result. Random background noise in data will look like a pattern given enough shuffling and combinations. It is by trying to reproduce findings, and publishing both the times a theory worked out and the times it failed, that we can see how solid a theory is overall. If positive findings have a higher probability of being published than negative ones, then literature becomes biased samples.

It is by trying to reproduce findings, and publishing both the times a theory worked out and the times it failed, that we can see how solid a theory is overall.

There is reason to be optimistic. A number of online initiatives have sprung up to in order to address these problems, like the Open Science Framework. Furthermore, some journals are starting to recognize the trouble with pushing for novel, pioneering research. While psychology's "vast graveyard of undead theories", as one behavioral scientist put it, seems to be psychology's most legitimate criticism, a number of caricatured criticisms are more well-known. Keith Stanovich addresses a large number of them in How to Think Straight About Psychology, and Scott Lilienfeld, Professor of Psychology at Emory University, has also published about the more popular criticisms. Those will not be explored in depth here. Two of them are "Psychology is just common sense" and "Psychology cannot predict things exactly". A moment's thought about these claims reveals their lack of substance. Critical thinkers understand a variety of problems with relying on "common sense" as the arbiter of truth, and behavior is inherently too complicated to predict. Yet with the power of statistical methodologies, psychological researchers do predict human behavior at a rate much greater than chance. Medical researchers also use these methods, and cognitive neuroscientists use them in conjunction with some of the same experimental paradigms as cognitive psychology.

Human thought and behavior are far more complicated than anything you can fit in a test tube. Yet researchers who are interested in questions about humans do the best they can to apply scientific thinking to behavior, and for that, they should be applauded. The problems affecting hypothesis falsification and reproducibility should not be de-emphasized, of course; even my introductory psychology professor lamented the lack of replication studies in his field toward the end of the semester. The problem of non-publication of data and non-replication is not uniform in psychology; some subfields are more affected by it than others. Some have pointed to clinical psychology as faring the worst and cognitive psychology as faring the best. Looking at the big picture, if academia does not incentivize null results or reproduced work, then surely other fields of scientific inquiry will be affected; psychology is not in a unique situation. It was only through psychological researchers publishing articles about these problems that we are aware of them. Self-criticism is the first step to self-correction. In fact, any field that exhibits these principles would be called scientific. After all, science is a process, a way of thinking. If someone exhibits this way of thinking and engages in the process, then that person is a scientist. A scientist is a scientist, regardless of what else we happen to call the person: sociologist, economist, archaeologist or, of course, psychologist.