

Generalization and the qualitative–quantitative debate

Geoff Norman¹

© Springer Science+Business Media B.V. 2017

We have a crisis on our hands. Many of us in health sciences education bridge a gap between two disciplines—psychology and medicine. We are now discovering that many of the findings that we viewed as robust and central to these disciplines are unreplicable (https://en.wikipedia.org/wiki/Replication_crisis).

The research in medicine is associated with one scientist—John Ioannidis, now at Stanford University. He has published several widely cited publications demonstrating that a number of ostensibly robust findings of clinical trials are not replicable (Ioannidis 2005a, b). In one study (2005b) he reviewed 45 highly cited studies; 13 had no replication studies, of the remainder, 20 (61%) replicated. While interesting and disturbing because of the real impact (positive or negative) of these interventions on health, the findings are likely of peripheral interest to educators.

However, lack of replicability in psychology is much closer to home. There have been a number of papers and articles raising issues about the extent to which results in psychology can be replicated. At the extreme, there have been several high profile exposures of scientific fraud. Perhaps the most egregious is Diederik Stapel, a social psychologist then at Tilburg, who made a lifetime career of fabricating data (https://en.wikipedia.org/wiki/Diederik_Stapel). At last count he has been forced to retract 58 publications. One has to hope that such practices are rare.

However a more prevalent occurrence is that findings that were thought to be robust cannot be reproduced. Some examples:

Dijksterhuis et al. (2006) achieved fame (and the inevitable Science article Dijksterhuis and Van Olden 2006) with something called the “Deliberation without attention” effect, which asserts that participants who put off a decision without

✉ Geoff Norman
norman@mcmaster.ca

¹ McMaster University, Hamilton, ON, Canada

consciously thinking about it are more satisfied down the line that either those who make an immediate decision or those who consciously dwell on the decision. There have been a number of failures to replicate, including one from medical education (Mamede, Schmidt, Rikers, Custers and Splinter 2010).

Bargh, Chen & Burrows (1996), Bargh & Chartrand (1999) did one better—he got an article in *Scientific American* (2014). He is famous for studies that are purported to show that people are unaware of the basis for their behaviours. As one classic example (1996), he gave undergraduate students lists of words in scrambled sentences. For one group, half the words were associated with the elderly (e.g. Florida, bald, wrinkled). They then walked down the hall to a second study room. Researchers inconspicuously timed how long they took to make the short walk. The group with “elderly” words took significantly longer, although on debriefing they had no conscious knowledge of the placing of “elderly words” or the change in their walking. Although other have had difficulty replicating the study, it was replicated in medical education. Hanson et al. (2007) studied a group of adolescents who had simulated depression and suicide for an OSCE. They were slower by a factor of 2.5 on a walk test immediately after the simulation. Two weeks later the difference in walk test was no longer significant.

One of Bargh’s harshest critics is Kahneman, the Nobel-prize-winning psychologist famous for his extensive research with Tversky on heuristics and biases. They have also made it into *Science* (Tversky and Kahneman 1974, 1981) as well as achieving popularity in medical education (Croskerry 2003). But they too have a few skeletons in their closet. Lopes (1991), in a little-known review, points out a number of areas where Tversky and have created very particular and atypical experimental situations in order to find their effects. Gigerenzer (1991) has shown that some of their biases are an artifact of the experimental method. Moreover, while in his best-selling book (2011), Kahneman claims unequivocally that these biases are hard-wired and cannot be unlearned, a number of studies have shown that supposedly robust effects like framing and base rate neglect are not or minimally present with medical experts (Christensen et al. 1995; Weber et al. 1993). And there is also a medical education connection. While the definitions of the heuristics in Kahneman’s hands appear unequivocal, Zwaan et al. (2017) showed that purported experts are completely unable to agree on the presence or absence of specific biases, and conversely are themselves strongly influenced by hindsight knowledge of the outcome

Qualitative researchers in our midst might well be feeling a bit justifiably smug at this point. After all, it is an axiom of their discipline that observations don’t generalize; every observation is so influenced by contextual details that replication is bound to fail. As Guba and Lincoln, in their iconic article (1994), state with reference to the constructivist view of knowledge:

Knowledge consists of those constructions about which there is relative consensus... among those competent to interpret the substance of their construction. Multiple “knowledges” can coexist when equally competent interpreters disagree, and/or depending on social, political, cultural, economic, ethnic and gender factors that differentiate the interpreters. These constructions are subject to continuous revision, with changes most likely to occur when relatively different constructions are brought into juxtaposition in a dialectical context.

Take a deep breath! I think what they're saying is that all knowledge is negotiable and depends on who is looking at it. It's interesting that my spellcheck redlined "knowledges". It's never heard of a plural for knowledge either. (I note in passing that it seems Donald Trump didn't invent alternative facts—Egon Guba did).

What does this mean? Their take home message is that knowledge resides in the individuals and contexts. Nothing can be expected to generalize. Indeed, even generalizations about the specific observations are suspect and relative. In short, their response to the finding that psychology studies don't replicate may well be a shrug of the shoulders, and a quick "So what did you expect?". In the field they work, where the questions they ask are frequently around the relation between individuals and the social and cultural context they are in, they're likely right. Things don't "replicate"; indeed replication is not anywhere in the vocabulary.

But there's a problem with this. Some things DO generalize. I came from physics, and I am absolutely certain that Newton's laws work just as well on the moon or Pluto as they do in New York. Similarly, some findings in psychology do replicate across widely different contexts, as we shall see in a moment. There are no multiple competing knowledges dependent on the actors in the sense that Guba and Lincoln (1994) describe.

As anyone who has engaged in the culture wars between qualitative and quantitative researchers will attest, the debate between the two groups are unlikely to resolve anytime soon. As the discussion above indicates, in the circles they move in, it's likely they are both right. But the irony in this whole debate is that, in their dogmatism, both the positivists and the constructivists are guilty of assuming that the context in which they work is a sound basis for generalizing about all of science. To put it bluntly, at the risk of offending some, constructivists are going around the world making sweeping generalizations about how you can't make sweeping generalizations. Conversely, as a previous editorial described (Norman 2016) many psychologists are guilty of devising materials that are uniquely designed to test their particular hypotheses, inflicting them on a barely willing sample of undergraduates who are almost like indentured labour, and then when they are done, making gross and sweeping generalizations about what "people" do under this or that circumstances (Lopes 1991).

Well it's one thing to say the truth lies somewhere in the middle; it's another to prove it. Recently, for me at least, the proof emerged from an unlikely quarter when a friend who is a retired physicist sent me a copy of an article in the Proceedings of the National Academy of Science that described how a number of psychologists have decided to do something about the charge that so many fundamental findings in psychology do not replicate, and started the Replication Project. This endeavour set out to systematically replicate 100 of the classic findings in psychology. The findings are disheartening; only 39% of the studies could be replicated. Analysis identified some of the "usual suspects"—low statistical power, small samples, small effect size. But a further analysis by van Bavel et al. (2016) took a completely different approach that led to real insights.

They reasoned that the lack of replication may be a result of "context effects". Taking a page from the constructivist hymn book, they state:

Many scientists have also argued that the failure to reproduce results might reflect contextual differences—often termed "hidden moderators"—between the original research and the replication attempt...

Understanding contextual influences on behavior is not usually considered an artifact or a nuisance variable but rather can be a driving force behind scientific inquiry and discovery.

However, rather than discussing context as a potential mediator, they went one better and operationalized it. Three reviewers examined each of the 100 findings, without knowing the results of the replication, and assessed “how likely the effect reported in the abstract of the original study was to vary by context—defined broadly as differing in time (e.g., pre vs. post-Recession), culture (e.g., individualistic vs. collectivistic culture), location (e.g., rural vs. urban setting), or population (e.g., a racially diverse population vs. a predominantly White population)” on a five point scale. They then conducted a regression analysis looking at whether or not the study did actually replicate to determine how much the rating of context affected the likelihood of replication. Context had a correlation of $-.23$ with likelihood of replication and was one of three significant predictors, along with statistical power and surprise of the result. A subsequent paper specifically examined the subdisciplines of social and cognitive psychology, and found that 28% of social psych findings replicate versus 53% of cognitive psychology, mainly because social psychology studies are more vulnerable to context effects.

These findings, as well as salvaging the reputation of cognitive psychology, at least to some degree, also pour some water on the longstanding raging debate between constructivists and positivists, qualitative and quantitative researchers. The evidence suggests that both are indeed correct in the domains in which they function. For the questions addressed by constructivists, findings will indeed be highly contextualized, severely limiting generalizability. Conversely, while it will likely remain the case that cognitive psychology will never achieve the universality of physics, many of the fundamental findings of the field can be assumed to be generalizable.

References

- Bargh, J. A. (2014). Our unconscious mind. *Scientific American*, *310*, 30–37.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244.
- Christensen, C., Heckerung, P., Mackesy-Amiti, M. E., Bernstein, L. M., & Elstein, A. S. (1995). Pervasiveness of framing effects among physicians and medical students. *Journal of Behavioral Decision Making*, *8*, 169–180.
- Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*, 775–780.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & Van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, *311*, 1005–1007.
- Dijksterhuis, A., & Van Olden, Z. (2006). On the benefits of thinking unconsciously: Unconscious thought can increase post-choice satisfaction. *Journal of Experimental Social Psychology*, *42*, 627–631.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology*, *2*, 83–115.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, *2*(163–194), 105.
- Hanson, M. D., Niec, A., Pietrantonio, A. M., Johnson, S., Young, M., High, B., et al. (2007). Effects associated with adolescent standardized patient simulation of depression and suicidal ideation. *Academic Medicine*, *82*, S61–S64.
- Ioannidis, J. P. (2005a). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.
- Ioannidis, J. P. (2005b). Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, *294*, 218–228.
- Kahneman, D. (2011). *Thinking, fast and slow*. Basingstoke: Macmillan.
- Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology*, *1*, 65–82.

- Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Splinter, T. A., & van Saase, J. L. (2010). Conscious thought beats deliberation without attention in diagnostic decision-making: At least when you are an expert. *Psychological Research, 74*, 586–592.
- Norman, G. R. (2016). A bridge too far. *Advances in Health Sciences Education, 21*, 251–256.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
- Van Bavel, J. J., Mende-Siedleckia, P., Brady, W. J., & Reinero, W. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Science, 23*, 6454–6459.
- Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1151.
- Zwaan, L., Monteiro, S., Sherbino, J., Ilgen, J., Howey, B., & Norman, G. (2017). Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Quality and Safety, 26*(2), 104–110.