



Metascience could rescue the 'replication crisis'

Independent replication of studies before publication may reveal sources of unreliable results, says Jonathan W. Schooler.

If you witness a bank robbery, what should you do? Tell the police what you saw? And then volunteer to identify the arrested suspect in a line-up? That sounds like the right response, but it might not be. Almost 25 years ago, a colleague and I published a psychology study (J. W. Schooler and T. Y. Engstler-Schooler *Cogn. Psychol.* 22, 36–71; 1990) that indicated that to describe the physical appearance of the criminal could make it harder for a witness to subsequently identify the person they saw.

The effect is called 'verbal overshadowing' and its discovery proved controversial, and not simply because of what it means for detectives. Over the years, other researchers (myself included) have had mixed success replicating the finding. Some have doubted whether it exists at all. That 1990 research project has been used as an example of the 'replication crisis' that has engulfed science in recent years. In disciplines such as medicine, psychology, genetics and biology, researchers have been confronted with results that are not as robust as they originally seemed.

In response, scientists have launched various replication projects to assess the robustness of published research. In psychology, numerous labs have volunteered to re-run studies, with the methods vetted by the original researchers. My discovery of the verbal-overshadowing effect was an obvious target for this approach, and so, last year, psychologists at 31 different laboratories across the world signed up to repeat the study and report the results.

How did I feel about having work scrutinized in this way? I thought that I could not lose. Positive replication would confirm the important verbal-overshadowing effect. Failure to replicate would be more evidence for the 'decline effect', an idea I endorse that the size of an effect decreases over repeated replications, for reasons that are not fully understood.

Unfortunately for a replication study, there was a mistake in the timing parameters of the initial experimental protocol. Still, this unexpected negative turn of events took a positive spin as the deviation from the original protocol, once identified and fixed, generated some useful comparative data.

Some 22 of the original 31 repeating labs went on to follow the corrected study protocol. Pooled, the results confirmed the original finding. The verbal-overshadowing effect was clearly demonstrated (although the effect size was smaller than in our 1990 research and observed only when the original parameters, in the corrected protocol, were followed).

The outcome is a genuine victory for the emerging field of metascience, an approach in which science turns the lens of scrutiny on itself. Metascience, the science of science, uses rigorous methods to examine how scientific practices

influence the validity of scientific conclusions. It has its roots in the philosophy of science and the study of scientific methods, but is distinguished from the former by a reliance on quantitative analysis, and from the latter by a broad focus on the general factors that contribute to the limitations and successes of research.

Large-scale replication efforts such as this one are important, but they are expensive, time-consuming and impracticable for the vast majority of scientific studies. Rather than focus exclusively on whether past studies stand up, we need a clearer sense of the processes that influence the reliability of new findings. These could include how 'invested' researchers are in the original hypothesis, the number of times a protocol is repeated and how the methods and outcomes are assessed and written up.

Together with labs at three other universities, my research group has begun an initiative to test the reproducibility of our science. Rather than re-examine published studies, each lab has agreed to allow the other three to generate new findings, replicate each others' experiments and compare the results of new studies before they are published. We will then be able to judge differences, for example, in the results obtained by the originating lab compared with those that follow the idea, and so, in theory, have less invested in the results.

It is clearly not feasible for all researchers to follow this approach in their routine work. But it should offer a valuable academic exercise to examine the factors that affect reproducibility as they arise during the course of research.

By pre-registering all aspects of new scientific studies and then repeatedly trying to replicate them, the project allows careful scrutiny of all parts of the research process, from inception to replication. If the studies replicate flawlessly, we will have established a gold standard for reproducible studies. If they do not, then our approach will present an opportunity to rigorously assess the reasons.

Some might suggest that the focus on replication within psychology is an indictment of the field. It is precisely the opposite. All fields face problems with reproducibility, and psychology should be applauded for its willingness to tackle the issue empirically.

In fact, psychological science has long been at the forefront of refining and improving the scientific process. The understanding of experimenter expectancy effects (a form of cognitive bias) and the importance of double-blind trials emerged first in psychology. Such self-examination can only strengthen the scientific process for all. ■

Jonathan W. Schooler is a psychologist at the University of California, Santa Barbara.
e-mail: jonathanwschooler@gmail.com

WE NEED A CLEARER SENSE OF THE PROCESSES THAT INFLUENCE THE RELIABILITY OF NEW FINDINGS.

➔ NATURE.COM
Discuss this article online at:
go.nature.com/8iabws