

Psychologists strike a blow for reproducibility

Thirty-six labs collaborate to check 13 earlier findings.

Ed Yong

26 November 2013

A large international group set up to test the reliability of psychology experiments has successfully reproduced the results of 10 out of 13 past experiments. The consortium also found that two effects could not be reproduced.

Psychology has been buffeted in recent years by mounting concern over the reliability of its results, after repeated failures to replicate classic studies. A failure to replicate could mean that the original study was flawed, the new experiment was poorly done or the effect under scrutiny varies between settings or groups of people.

To tackle this 'replicability crisis', 36 research groups formed the Many Labs Replication Project to repeat 13 psychological studies. The consortium combined tests from earlier experiments into a single questionnaire — meant to take 15 minutes to complete — and delivered it to 6,344 volunteers from 12 countries.

The team chose a mix of effects that represent the diversity of psychological science, from classic experiments that have been repeatedly replicated to contemporary ones that have not.

Ten of the effects were consistently replicated across different samples. These included classic results from economics Nobel laureate and psychologist Daniel Kahneman at Princeton University in New Jersey, such as gain-versus-loss framing, in which people are more prepared to take risks to avoid losses, rather than make gains¹; and anchoring, an effect in which the first piece of information a person receives can introduce bias to later decisions². The team even showed that anchoring is substantially more powerful than Kahneman's original study suggested.

Encouraging outcomes

"This is a really important initiative for psychology," says Danny Oppenheimer, a psychologist at the University of California, Los Angeles, whose work was under scrutiny but who did not take part in the collaboration. "It means that the replicability problem, while by no means trivial, may not be as widespread as some critics of the field have suggested."

Project co-leader Brian Nosek, a psychologist at the Center of Open Science in Charlottesville, Virginia, finds the outcomes encouraging. "It demonstrates that there are important effects in our field that are replicable, and consistently so," he says. "But that doesn't mean that 10 out of every 13 effects will replicate."

Kahneman agrees. The study "appears to be extremely well done and entirely convincing", he says, "although it is surely too early to draw extreme conclusions about entire fields of research from this single effort". [Kahneman published an open letter](#) in 2012 calling for a "daisy chain" of replications of studies on priming effects, in which subtle, subconscious cues can supposedly affect later behaviour.

Of the 13 effects under scrutiny in the latest investigation, one was only weakly supported, and two were not replicated at all. Both irreproducible effects involved social priming. In one of these, people had increased their endorsement of a current social system after being exposed to money³. In the other, Americans had espoused more-conservative values after seeing a US flag⁴.

Social psychologist Travis Carter of Colby College in Waterville, Maine, who led the original flag-priming study, says that he is disappointed but trusts Nosek's team wholeheartedly, although he wants to review their data before commenting further. Behavioural scientist Eugene Caruso at the University of Chicago in Illinois, who led the original currency-priming study, says, "We should use this lack of replication to update our beliefs about the reliability and generalizability of this effect", given the "vastly larger and more diverse sample" of the Many Labs project. Both researchers praised the initiative.

Open documentation

The Many Labs team, which was also coordinated by Richard Klein and Kate Ratliff from the University of Florida in Gainesville and Michelangelo Vianello from the University of Padua, Italy, found that the results were largely unaffected by the nationality of the

volunteers or the setting of the experiments — whether delivered online or in a lab setting. When there was variation, it was limited to the large and obvious effects such as anchoring rather than the small and subtle ones such as being primed by seeing a flag.

This contradicts the frequently cited idea that some psychological studies, especially those on subtle social-priming effects, are hard to replicate because they are sensitive to factors such as the sample being studied or the skill of the original investigators. The fact that social priming studies have been hard to replicate “has been difficult for me personally,” says Nosek, “because social priming is an area that’s important for my research”.

The plan for the Many Labs project was vetted by the original authors where possible, was documented openly, and was registered with the journal *Social Psychology* and its methods were peer-reviewed before any experiments were done. The results have now been submitted to the journal and are [available online](#). “That sort of openness should be the standard for all research,” says Daniel Simons of the University of Illinois at Urbana–Champaign, who is coordinating a similar collaborative attempt to verify a classic psychological effect not covered in the present study. “I hope this will become a standard approach in psychology.”

Oppenheimer says that other disciplines could benefit from Many Labs' approach. “Psychology isn't the only field that has had issues with replication in recent years.”

Nature | doi:10.1038/nature.2013.14232

References

1. Tversky, A. & Kahneman, D. *Science* **211**, 453–458 (1981).
2. Jacowitz, K. E. & Kahneman, D. *Pers. Soc. Psych. Bull.* **21**, 1161–1166 (1995).
3. Caruso, E. M., Vohs, K. D., Baxter, B. & Waytz, A. J. *Exp. Psych: Gen.* **142**, 301–306 (2013).
4. Carter, T. J., Ferguson, M. J. & Hassin, R. R. *Psych. Sci.* **22**, 1011–1018 (2011).