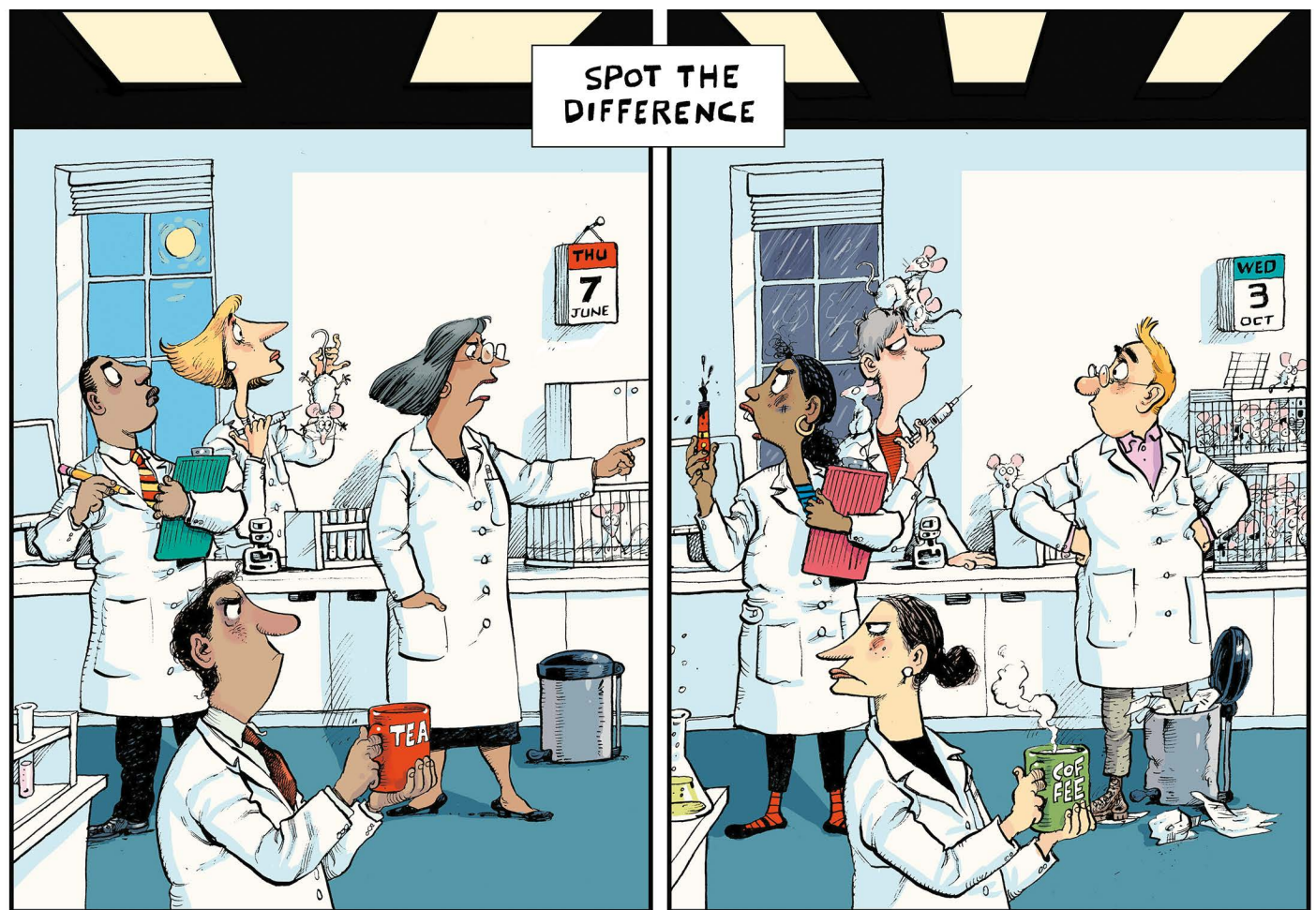


Comment



Argue about what a replication means before you do it

Brian A. Nosek & Timothy M. Errington

To avoid stalemates and provide lessons, replicators and original researchers must reach agreement on a study design and set out expectations ahead of time.

The two of us have spent years coordinating replications of published studies. The most consistent outcomes are confusion and disagreement, particularly when outcomes seem to contradict original findings.

We saw this in the Reproducibility Project: Cancer Biology, in which we managed attempts to replicate experiments from high-impact papers¹. Among the 50 replication experiments completed (from 23 papers), one required transplanting leukaemia cells into immunocompromised mice and letting the cells grow before administering a potential treatment.

Neither our team, the reviewers nor the original authors thought that the immunosuppression technique we proposed, which differed from that in the original study, would affect the key question of whether the treatment lengthened survival². Until we did the replication and observed no difference in survival³. Then, the reviewers said that this difference was crucial.

Similar scenarios play out in other disciplines. When one of us (B.A.N.) worked as a guest editor for the journal *Social Psychology*, a reviewer objected to the setting of a replication study selected to test the hypothesis that Asian women perform better on a mathematics

test when reminded of their Asian identity. A second replication study was then commissioned in a setting that met the reviewer's criteria. The replication failed in the second, commissioned setting but mostly succeeded in the original setting that had been deemed inappropriate by the reviewer^{4,5}. After learning the results, the reviewer then stated that the second study was obviously flawed.

Failure to replicate often brings intellectual gridlock. Some researchers insist that a replication refutes the original paper's ideas; others find flaws in the reproduced work. Both replicators and original authors defend their conclusions – or at least their competence – rather than getting on with the difficult, intellectual work of using new evidence to revise ideas. Human nature and the academic incentive system make it hard to do otherwise.

How can researchers avoid such stalemates? We need to spend more time early on resolving what is to be tested, the crucial features for doing so and the insight we expect. We need a process that appeals to our better natures, or at least requires that we reveal our lesser selves. The approach should favour seeking an accurate answer over defending previous results.

We call it precommitment. After a paper is made public, but before it is replicated, the original authors and independent replicators collaborate to design a replication experiment that both agree will be meaningful, whatever the results. This process will be documented using preregistration or, ideally, a Registered Report (see 'Routes to replication').

Solid scaffolds

Over the past decade or so, philanthropists, government funders and journal editors have started to devote policies, programmes and money to replication research. Methodologists are working out how replication can sift promising paths from dead ends.

But we still lack the tools to make the most of replications. Too often, they are seen as a hostile act rather than an ordinary and desirable part of the scientific process. Precommitment will make replications more informative. It will favour collaboration over confrontation and promote scientific humility. It will focus energy on improving the quality of replications, on maximizing what can be learned from them, and on reducing the tendency to assess replications depending on whether they fit previous beliefs.

Replication depends on creating the conditions necessary to repeat a finding. Even if two studies use the same protocol, they will still differ in innumerable ways: time

of year, climate, samples used and identities of the experimenters, to name a few. Conducting a replication demands a theoretical commitment to the features that matter⁶.

It is difficult to make that commitment after results are known. Ask whether temperature (or age or light or language or any other variable) matters before a replication is done, and many researchers will acknowledge that they don't know or that they didn't even realize that condition had been held constant in the original experiments. Ask the same question after a replication has failed, and hindsight bias supplies a different answer. "Of course, temperature [or age or light or language or any other variable] matters." A replication study cannot test current understanding if any outcome can be accommodated by adding assumptions after the fact.

Worth a gamble

What, then, constitutes a theoretical commitment? Here's an idea from economists: a theoretical commitment is something you're willing to bet on.

"The approach should favour seeking an accurate answer over defending results."

If researchers are willing to bet on a replication with wide variation in experimental details, that indicates their confidence that a phenomenon is generalizable and robust. Those who will only precommit after narrowing the conditions shrink the phenomenon to match their confidence.

Let's say a study reports that regular exercise improves memory. Proponents might insist on reusing methods from the original study, such as limiting the definition of 'regular exercise' to running not cycling. They might also insist on narrower conditions than were originally specified – perhaps that the memory test be administered only at night, among adults under 35 and in the United Kingdom. This insistence suggests that the proponents lack full confidence in the general claim that regular exercise improves memory, and actually believe a much narrower claim. If they cannot suggest any design that they would bet on, perhaps they don't even believe that the original finding is replicable.

Revise ideas

A good-faith replication study matches its design to the original claims. If claims specify location, the replication must take location

Collaborative adversaries

Proponents and sceptics should 'precommit' to an agreed replication design and make their differing predictions explicit.

Precommitment draws on adversarial collaboration, an approach put forward in the early 2000s by Nobel-prizewinning psychologist Daniel Kahneman and his colleagues (see, for example, ref. 11). In this process, researchers with incompatible views agree on and conduct tests to help settle a debate with the help of an arbiter.

In 2015, researchers who disagreed about how a type of eye movement (called saccades) affected memory recall combined adversarial collaboration with preregistration of the study design and of their expectations. All authors and the mediator then reported their findings and disparate analyses in a single, joint paper¹².

Key to the success of the approach is having adversaries who are good listeners and who genuinely want to work out what the opposing side contends. So says Dawid Potgieter, who has spearheaded adversarial collaboration efforts through his role as director of programmes in discovery science at the Templeton World Charity Foundation in Nassau, the Bahamas.

Similar advice comes from an international group of researchers that is testing five opposing models for processes that underlie cooperation. They advocate aligning researchers' mindsets before work starts, such as by framing joint goals, fostering curiosity and assuming competence and good intentions¹³.

Precommitment would take these ideals and pin a concrete process on individual papers: an arbitrary but expedient and manageable unit. Learning to appraise experiments and to evaluate others' critiques of your work is central to robust science. We should make it part of the scientific infrastructure.

Comment

into account. If the original claims ignore or dismiss age, then the replication need not consider it. Designing a replication study with input from proponents, sceptics and neutral bystanders can clarify the boundaries of claims, especially ones that went unspecified.

We have witnessed too many debates that stalled because proponents and sceptics misunderstood or talked past one another. To fix this, we need an efficient process to manage replication designs a priori. It must produce claims that are sufficiently well articulated to be tested, and needs to cope with the distrust that arises between people who disagree.

The most informative replication will occur when proponents and sceptics each endorse the research design but predict different results. Examples of adversarial collaboration illustrate both the difficulty and the potential of this approach (see ‘Collaborative adversaries’). For instance, leading neuroscientists with different perspectives on consciousness came together in a project initiated by the Templeton World Charity Foundation, based in Nassau in the Bahamas, to design experiments for which their theories demanded different outcomes. People got worked up. There was shouting. But, after two days, experiments were proposed. The results, which should be available later this year, won’t settle the debates about consciousness, but they should advance understanding (see go.nature.com/3gqou5u).

Make it happen

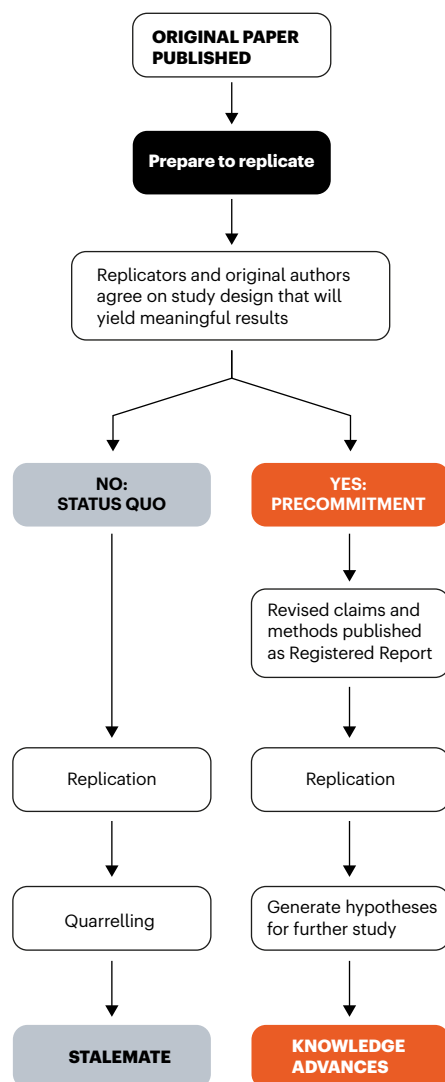
Seven years ago, we conducted an exercise that convinced us how the principle of precommitment might work. Now, the basic infrastructure to implement precommitment is widely available: Registered Reports. In this system, authors, reviewers and editors evaluate a study before it is performed. Assuming the research question is important and the methodology is of high quality, that work is accepted for publication before results are known⁷.

Our proof-of-concept exercise encompassed 15 replication papers published as a special issue of *Social Psychology*⁸. Teams proposed replication studies of important findings in the field. Original authors and other expert reviewers critiqued the proposed methodology. Although exchanges were occasionally tense and confrontational, these ‘adversarial teams’ and journal editors worked on a shared goal: designing a methodology that would make the replication results meaningful. That still didn’t eliminate all controversy – far from it. Indeed, after publication, one paper spurred what came to be called ‘repligate’ involving name-calling, competing reanalyses and reflections on civility (see go.nature.com/3fitemmf).

The other papers better illustrate the desired outcome: proponents and sceptics observed the findings, debated their meaning and

ROUTES TO REPLICATION

Precommitment rewards authors for providing clear, testable claims and helps to advance knowledge.



offered alternative explanations. Because the methodology and analyses had already been agreed, alternative explanations for the replication results were framed appropriately as areas for follow-up research, rather than as a necessary component for a valid experiment.

For example, one paper that failed to replicate findings that superstitions can improve performance discussed the possibility that the original finding was a false positive or that the particular type of task or belief could account for the differences⁹. Crucially, these potential moderating influences were described as hypotheses for future study, not as explanations for unexpected outcomes. In many ways, the process felt like the ideal of how we imagined science to operate when we entered the field.

The Registered Reports format is now offered by more than 250 journals. *Nature Communications* became one of the latest of these earlier this month. Funders such as the US-based Flu Lab and the Children’s Tumor Foundation in New York City have each

partnered with the scientific publisher PLOS to fund Registered Reports of important findings in their fields. These journals (and many others) archive accepted Registered Reports on a platform (<http://osf.io/rr>) that is maintained by the non-profit Center for Open Science, where we work. Our centre can support journals that implement a precommitment process combined with Registered Reports.

Changing mindsets

Researchers see findings as personal possessions, so a replication puts them at risk of losing something and amplifies their desire to be right. Precommitment offers an opportunity for everyone, proponent or sceptic, to shift away from that unproductive framework and pursue a shared goal of getting it right. Original authors can be rewarded for generating clear and specific testable claims, and for transparently reporting how to test them.

We think that the visibility of precommitments will be sufficient to shift incentives. After all, which scientist would you admire more – one who never agrees with independent tests of their findings, or one who willingly precommits and revises beliefs when new results suggest they were wrong? We have evidence that researchers and the public prefer the latter¹⁰. Critics will counter that some experiments are inherently messy or require arcane techniques, or that replicators’ energies would be better spent on original ideas. We argue that testing existing claims to improve understanding is essential for progress.

Eventually, precommitment should become an expectation. Whether or not the results come out as proponents or sceptics expected, knowledge comes out ahead.

The authors

Brian A. Nosek is co-founder and executive director of the Center for Open Science, Charlottesville, Virginia; and professor of psychology at the University of Virginia, Charlottesville, USA. **Timothy M. Errington** is director of research at the Center for Open Science, Charlottesville, Virginia, USA. e-mails: nosek@cos.io; tim@cos.io

- Errington, T. M. et al. *eLife* **3**, e04333 (2014).
- Fung, J. J. et al. *eLife* **4**, e08997 (2015).
- Shan, X., Fung, J. J. & Kosaka, A. *eLife* **6**, e25306 (2017).
- Gibson, C. E., Losee, J. & Vitellio, C. *Soc. Psychol.* **45**, 194–198 (2014).
- Moon, A. & Roeder, S. *Soc. Psychol.* **45**, 199–201 (2014).
- Nosek, B. A. & Errington, T. M. *PLoS Biol.* **18**, e3000691 (2020).
- Chambers, C. *Nature* **573**, 187–189 (2019).
- Nosek, B. A. & Lakens, D. *Soc. Psychol.* **45**, 137–141 (2014).
- Calin-Jageman, R. J. & Caldwell, T. L. *Soc. Psychol.* **45**, 239–245 (2014).
- Ebersole, C. R., Axt, J. R. & Nosek, B. A. *PLoS Biol.* **14**, e1002460 (2016).
- Kahneman, D. *Am. Psychol.* **58**, 723–730 (2003).
- Matzke, D. et al. *J. Exp. Psychol. Gen.* **144**, e1–e15 (2015).
- Ellemers, N., Fiske, S. T., Abele, A. E., Koch, A. & Yzerbyt, V. *Proc. Natl Acad. Sci. USA* **117**, 7561–7567 (2020).