

strument flown to another planet, used to analyze the crystal structure of samples—would find no place on board.

That hasn't stopped scientists from proposing even more complicated experiments. In January, NASA received 58 instrument proposals. One would seek to germinate plants as a first step toward increasing the amount of oxygen on the planet. Another team would build a solar-powered helicopter that could fly reconnaissance sorties of hundreds of meters a day.

Payload selection is expected in mid-July. Jack Mustard, a planetary scientist at Brown University who led the science definition team, hopes NASA will opt for a conservative payload—something more along the lines of what is riding aboard the small Opportunity rover, now in its 10th year of operation. The same considerations will drive the choice of a landing site. Sites that require more than 10 kilometers of driving will be problematic, Mustard says. Sites where sample collection can begin immediately after landing are “going to trump.”

Chen, taking over for Steltzner in leading the entry, descent, and landing phase of the mission, hopes he can help by delivering the rover to just the right spot. He is considering two modifications to the Curiosity system that could improve landing precision. One is a device that would deploy the entry capsule's parachute based on distance to the target, rather than an estimate of velocity. This range trigger would shrink the landing ellipse—where engineers are confident the rover will end up—from 25 kilometers to 13 kilometers long. According to project manager John McNamee, the trigger would require changing just a few lines of code.

Another change, called terrain relative navigation (TRN), would allow the rover to land at sites that would otherwise be too hazardous. It would outfit the descent module to compare real-time images of the terrain it was approaching with stored images of the landing site, enabling the sky crane to avoid regions with too many rocks. McNamee says the price tag for the TRN is higher—on the order of \$10 million—and could ripple through the landing system and affect it in other ways. He wants to be convinced that the change is absolutely necessary before allowing it.

Scientists at the workshop spoke with near unanimity, however: The TRN would open up the most interesting landing sites, which previous missions had discarded for safety reasons. Furthermore, the technology would help guide the subsequent missions needed to fetch the cache. Steltzner concurs. “We all kinda think that the TRN is the steering wheel, the headlights,” he says. ■



A 2008 study (right) showing that cleanliness influences moral judgments was not replicated in a new study (left).

PSYCHOLOGY

Replication effort provokes praise—and ‘bullying’ charges

Global network fails to confirm 10 of 27 psychology findings, but some call project an inquisition

By John Bohannon

After a string of scandals involving accusations of misconduct and retracted papers, social psychology is engaged in intense self-examination—and the process is turning out to be painful. This week, a global network of nearly 100 researchers unveiled the results of an effort to replicate 27 well-known studies in the field. In more than a third of the cases, the result was a complete failure.

As the replicators see it, the failed do-overs are a healthy corrective. “Replication helps us make sure what we think is true really is true,” says Brent Donnellan, a psychologist at Michigan State University in East Lansing who has undertaken three recent replications of studies from other groups—all of which came out negative. “We are moving forward as a science,” he says.

But rather than a renaissance, some researchers on the receiving end of this organized replication effort see an inquisition. “I feel like a criminal suspect who has no right to a defense and there is no way to

win,” says psychologist Simone Schnall of the University of Cambridge in the United Kingdom, who studies embodied cognition, the idea that the mind is unconsciously shaped by bodily movement and the surrounding environment. Schnall’s 2008 study finding that hand-washing reduced the severity of moral judgment was one of those Donnellan could not replicate.

About half of the replications are the work of Many Labs, a network of about 50 psychologists around the world. The results of their first 13 replications, released online in November, were greeted with a collective sigh of relief: Only two failed. Meanwhile, Many Labs participant Brian Nosek, a psychologist at the University of Virginia in Charlottesville, put out a call for proposals for more replication studies. After 40 rolled in, he and Daniël Lakens, a psychologist at Eindhoven University of Technology in the Netherlands, chose another 14 to repeat.

The output of the new batch of replications, published alongside the previous 13 this week in an issue of *Social Psychology*

guest-edited by Nosek and Lakens, is less reassuring. All told, the researchers failed to confirm the results of 10 well-known studies, such as the social psychological effects of washing one's hands, holding cups of warm or cold liquid, or writing down flattering things about oneself. In another five cases, the replications found a smaller effect than the original study did or encountered statistical complications it did not report. For embodied cognition and also for behavior priming—the study of how exposure to one stimulus, such as the word “dog,” changes one's reaction to another, such as a photo of a cat—the results are particularly grim. Seven of the replications focused on experiments in these areas, and all but one failed.

No one is suggesting misconduct in any of the original studies, but the results are further blows to a field shaken several years ago when a towering figure in priming research, Diederik Stapel, confessed to faking data (*Science*, 7 December 2012, p. 1270). And earlier this month, Jens Förster of the University of Amsterdam, a pioneer of embodied cognition research, was accused by a Dutch government-appointed ethics panel of data manipulation—charges he denies (*Science*, 9 May, p. 566).

Nor should the results be taken as a general indictment of psychological research, because the targeted studies were not a random sample, Nosek says. “They are entirely cherry-picked,” he says, based on the importance of the original study and the feasibility of replicating it.

Some of the authors of the targeted studies, however, feel not just singled out but persecuted. Schnall, for example, contends that the replications were not held to the same peer-review standard as her original studies. “I stand by my methods and my findings and have nothing to hide,” she says.

The replications did employ an alternative model of peer review, called pre-registration, promoted by the Center for Open Science, a nonprofit organization co-founded by Nosek (*Science*, 30 March 2012, p. 1558). Before any data were collected, the replicators submitted their experimental design and data analysis plan to external peer reviewers, including the principal investigator of the original study. The subsequent data analysis and conclusions were reviewed only by Nosek or Lakens.

Schnall contends that Donnellan's effort was flawed by a “ceiling effect” that, essentially, discounted subjects' most severe moral sentiments. “We tried a number of strategies to deal with her ceiling effect concern,” Donnellan counters, “but it did not change the conclusions.” Donnellan and his supporters say that Schnall simply

tested too few people to avoid a false positive result. (A colleague of Schnall's, Oliver Genschow, a psychologist at Ghent University in Belgium, told *Science* in an e-mail that he has successfully replicated Schnall's study and plans to publish it.)

Some replicators leaked news of their findings online, long before publication and in dismissive terms. On his personal blog, Donnellan described his effort to repeat Schnall's research as an “epic fail” in a December post titled “Go Big or Go Home,” which was then widely circulated on Twitter. Donnellan defends the early announcement. “I feel badly, but the results are the results,” he says.

Schnall, however, says that her work was “defamed.” She believes she was denied a large grant in part because of suspicions about her work and says that a reviewer of one of her recently submitted papers “raised

spective, it's apparent that we can always learn something from a carefully designed and executed study.” Caruso now has a larger and more nuanced version of his study under way.

The replications in psychology reflect a growing trend in science (see table). The field's bruising experience shows that such efforts should be handled carefully, stresses Daniel Kahneman, a psychologist at Princeton University, whose work was successfully replicated by the Many Labs team. “The relationship between authors and skeptics who doubt their findings is bound to be fraught,” he says. “It can be managed professionally if the rules that apply to both sides are clearly laid out.”

To reduce professional damage, Kahneman calls for a “replication etiquette,” which he describes in a commentary published with the replications in

Repeat after me

Select efforts in replication of research

EFFORT	REPLICATION TARGET
Reproducibility Project: Cancer Biology	50 high-impact cancer studies published from 2010 to 2012
Reproducibility Project: Psychology	Articles published in 2008 from three psychology journals
Reproducibility Initiative	Hub for authors to request independent replications of their experiments
Many Labs project	Global network for orchestrating large replications
Reproducibility in Computer Science	Checks software code in 613 applied computer science papers
Crowdstorming project	More than 50 analysts address same research question using shared data set

the issue about a ‘failed’ replication.” She adds that her graduate students “are worried about publishing their work out of fear that data detectives might come after them and try to find something wrong.”

Other researchers whose work was targeted and failed to replicate told *Science* that they have had experiences similar to Schnall's. They all requested anonymity, for fear of what some in the field are calling “replication bullying.”

Yet some whose findings did not hold up are putting a positive spin on the experience. “This was certainly disappointing at a personal level,” says Eugene Caruso, a psychologist at the University of Chicago Booth School of Business in Illinois, who in 2013 reported a priming effect—exposing people to the sight of money made them more accepting of societal norms—that failed to replicate. “But when I take a broader per-

Social Psychology. For example, he says, “the original authors of papers should be actively involved in replication efforts” and “a demonstrable good-faith effort to achieve the collaboration of the original authors should be a requirement for publishing replications.” In the case of this week's replications, “the consultations did not reach the level of author involvement that I recommend.” However, he notes that “authors of low-powered studies with spectacular effects should not wait for hostile replications: They should get in front of the problem by replicating their own work.”

For his part, Nosek hopes that the tensions will be short-lived growing pains as psychology adjusts to a demand, from within and outside the field, for greater accountability. “Our primary aim is to make replication entirely ordinary,” he says, “and move it from a threat to a compliment.” ■