

MARCH 6 2016 8:02 PM

# Everything Is Crumbling

An influential psychological theory, borne out in hundreds of experiments, may have just been debunked. How can so many scientists have been so wrong?

By Daniel Engber



Lisa Larson-Walker

Nearly 20 years ago, psychologists Roy Baumeister and Dianne Tice, a married couple at Case Western Reserve University, devised a foundational experiment on self-control. “Chocolate chip cookies were baked in the room in a small oven,” they wrote in **a paper that has been cited more than 3,000 times**. “As a result, the laboratory was filled with the delicious aroma of fresh chocolate and baking.”



DANIEL ENGBER

In the history of psychology, there has never been a more important chocolate-y aroma.

Advertisement

Here's how that experiment worked. Baumeister and Tice stacked their fresh-baked cookies on a plate, beside a bowl of red and white radishes, and brought in a parade of student volunteers. They told some of the students to hang out for a while unattended, eating only from the bowl of radishes, while another group ate only cookies. Afterward, each volunteer tried to solve a puzzle, one that was designed to be impossible to complete.

Baumeister and Tice timed the students in the puzzle task, to see how long it took them to give up. They found that the ones who'd eaten chocolate chip cookies kept working on the puzzle for 19 minutes, on average—about as long as people in a control condition who hadn't snacked at all. The group of kids who noshed on radishes flubbed the puzzle test. They lasted just eight minutes before they quit in frustration.

The authors called this effect “ego depletion” and said it revealed a fundamental fact about the human mind: We all have a limited supply of willpower, and it decreases with overuse. Eating a radish when you're surrounded by fresh-baked cookies represents an epic feat of self-denial, and one that really wears you out. Willpower, argued Baumeister and Tice, draws down mental energy—it's a muscle that can be exercised to exhaustion.

That simple idea—perhaps intuitive for nonscientists, but revolutionary in the field—turned into a research juggernaut. In the years that followed, Baumeister and Tice's lab, as well as dozens of others, published scores of studies using similar procedures. First, the scientists would deplete subjects' willpower with a task that requires self-control: don't eat chocolate chip cookies, watch this sad movie but don't react at all. Then, a few minutes later, they'd test them with a puzzle, a game, or something else that requires mental effort.

Advertisement

Psychologists discovered that lots of different tasks could drain a person's energy and leave them cognitively depleted. Poverty-stricken day laborers in rural India might wear themselves out simply by **deciding whether to purchase a bar of soap**. Dogs might waste

their willpower by **holding back from eating chow**. White people might lose mental strength when they tried to **talk about racial politics with a black scientist**. In 2010, a group of researchers led by Martin Hagger put out a meta-analysis of the field—a study of published studies—to find out whether this sort of research could be trusted. Using data from **83 studies and 198 separate experiments**, Hagger’s team confirmed the main result. “Ego depletion” seemed to be a real and reliable phenomenon.

In 2011, Baumeister and John Tierney of the *New York Times* published a **science-cum-self-help book** based around this research. Their best-seller, ***Willpower: Rediscovering the Greatest Human Strength***, advised readers on how **the science of ego depletion could be put to use**. A glass of lemonade that’s been sweetened with real sugar, they said, could help replenish someone’s inner store of self-control. And if willpower works like a muscle, then regular exercise could boost its strength. **You could literally build character**, Baumeister said in an interview with the **Templeton Foundation**, a religiously inclined science-funding organization that has given him about **\$1 million** in grants. By that point, he told the *Atlantic*, the effects that he’d first begun to study in the late 1990s were established fact: “They’ve been replicated and extended in many different laboratories, so **I am confident they are real**,” he said.

But that story is about to change. A paper now in press, and due to publish next month in the journal *Perspectives on Psychological Science*, describes a massive effort to reproduce the main effect that underlies this work. Comprising more than 2,000 subjects tested at two-dozen different labs on several continents, the study found exactly nothing. **A zero-effect for ego depletion**: No sign that the human will works as it’s been described, or that these hundreds of studies amount to very much at all.

This isn’t the first time that an idea in psychology has been challenged—not by a long shot. A “**reproducibility crisis**” in psychology, and in **many other fields**, has now been well-established. A study out last summer tried to **replicate 100 psychology experiments** one-for-one and found that just 40 percent of those replications were successful. A **critique of that study** just appeared last week, claiming that the original authors made statistical errors—but that critique has itself been **attacked** for **misconstruing facts, ignoring evidence**, and indulging in some **wishful thinking**.

Advertisement

For scientists and science journalists, this **back and forth** is worrying. We'd like to think that a published study has more than even odds of being true. The new study of ego depletion has much higher stakes: Instead of warning us that any single piece of research might be unreliable, the new paper casts a shadow on a fully-formed *research literature*. Or, to put it another way: It takes aim not at the single paper but at the Big Idea.

Baumeister's theory of willpower, and his clever means of testing it, have been borne out again and again in empirical studies. The effect has been recreated in hundreds of different ways, and the underlying concept has been verified via meta-analysis. It's not some crazy new idea, wobbling on a pile of flimsy data; it's a sturdy edifice of knowledge, built over many years from solid bricks.

And yet, it now appears that ego depletion could be completely bogus, that its foundation might be made of rotted-out materials. That means an entire field of study—and significant portions of certain scientists' careers—could be resting on a false premise. If something this well-established could fall apart, then what's next? That's not just worrying. It's terrifying.

\* \* \*

#### Advertisement

Evan Carter was among the first to spot some weaknesses in the ego depletion literature. As a graduate student at the University of Miami, Carter set out to recreate the **lemonade effect**, first described in 2007, whereby the consumption of a sugary drink staves off the loss of willpower. "I was collecting as many subjects as I could, and we ended up having one of the largest samples in the ego-depletion literature," Carter told me. But for all his efforts, he couldn't make the study work. "I figured that I had gotten some bad intel on how to do these experiments," he said.

To figure out what went wrong, Carter reviewed the 2010 meta-analysis—the study using data from **83 studies and 198 experiments**. The closer he looked at the paper, though, the less he believed in its conclusions. First, the meta-analysis included only published studies, which meant the data would be subject to a standard bias in favor of positive results. Second, it included studies with contradictory or counterintuitive measures of self-control. One study, for example, suggested that depleted subjects would give more money to charity while another said depleted subjects would spend less time helping a stranger. When he and his adviser, Michael McCullough, reanalyzed the 2010 paper's data using state-of-the-art analytic methods, **they found no effect**. For a second paper published last year, Carter and

McCullough completed a second meta-analysis that included different studies, including 48 experiments that had never been published. Again, they found “**very little evidence**” of a real effect.

“All of a sudden it felt like everything was crumbling,” says Carter, now 31 years old and not yet in a tenure-track position. “I basically lost my compass. Normally I could say, all right there have been 100 published studies on this, so I can feel good about it, I can feel confident. And then that just went away.”

Not everyone believed Carter and McCullough’s reappraisal of the field. The fancy methods they used to correct for publication bias were new, and not yet fully tested. Several prominent researchers in the field **called their findings premature**.

Advertisement

But by this point, there were other signs of problems in the literature. The lemonade effect, for one, seemed **implausible on its face**: There’s no way the brain could use enough glucose, and so quickly, that drinking a glass of lemonade would make a difference. What’s more, several labs were able to produce the same result—restoration of self-control—by having people **swish the lemonade around their mouths** and spit it out instead of drinking it. Other labs discovered that a subject’s **beliefs and mindset** could also affect whether and how her willpower was depleted.

These criticisms weren’t fatal in themselves. It could be that willpower is a finite resource, but one that **we expend according to our motivations**. After all, that’s how money works: A person’s buying habits might encompass lots of different factors, including how much cash she’s holding and how she feels about her finances. But given these larger questions about the nature of willpower as well as the meta-analysis debate, the whole body of research began to seem suspicious.





Lisa Larson-Walker

In October 2014, the Association for Psychological Science **announced** it would try to resolve some of this uncertainty. APS would create a “**Registered Replication Report**”—a planned-out set of experiments, conducted by many different labs, in the hopes of testing a single study that represents an important research idea. Martin Hagger, who wrote the original 2010 meta-analysis, would serve as lead author on the project. Roy Baumeister would consult on methodology.

The replication team had to choose the specific form of its experiment: Which of the hundreds of ego-depletion studies would they try to replicate? Baumeister suggested some of his favorite experimental designs, but most turned out to be unworkable. The replication team needed tasks that could be reliably repeated in many different labs. The chocolate-chip-cookie experiment, for example, would never work. What if one lab burned the cookies? That would ruin everything!

Advertisement

With Baumeister’s counsel, Hagger’s team settled on a **2014 paper** from researchers at the University of Michigan. That study used a standard **self-control task**. Subjects watched as simple words flashed on a screen: *level*, *trouble*, *plastic*, *business*, and so on. They were asked to hit a key if the word contained the letter e, but only if it was not within two spaces of another vowel (i.e., they had to hit the key for *trouble* but withhold their button-press for

level and *business*). In the original study, this exercise of self-control produced a strong depletion effect. The subjects performed markedly worse on a **follow-up test**, also done on the computer.

The replication team ran that same experiment at **24 different labs**, including ones that translated the letter e task into Dutch, German, French, and Indonesian. Just two of the research groups produced a significant, positive effect, says study co-author Michael Inzlicht of the University of Toronto. (One appeared to find a negative effect, a reverse-depletion.) Taken all together, the experiments showed no signs whatsoever of Baumeister and Tice's original effect.

What, exactly, does that mean? At the very least, it tells us that one specific task—the letter e game—doesn't sap a subject's willpower, or else that the follow-up test did not adequately measure that depletion. Indeed, that's how Baumeister himself sees the project. "I feel bad that people went through all this work all over the world and did this study and found a whole bunch of nothing," he told me earlier this week, in a phone call from Australia. He still believes ego depletion is real. The tasks had failed, not the Big Idea.

In his lab, Baumeister told me, the letter e task would have been handled differently. First, he'd train his subjects to pick out *all* the words containing e, until that became an ingrained habit. Only then would he add the second rule, about ignoring words with e's and nearby vowels. That version of the task requires much more self-control, he says.

Second, he'd have his subjects do the task with pen and paper, instead of on a computer. It might take more self-control, he suggested, to withhold a gross movement of the arm than to stifle a tap of the finger on a keyboard.

If the replication showed us anything, Baumeister says, it's that the field has gotten hung up on computer-based investigations. "In the olden days there was a craft to running an experiment. You worked with people, and got them into the right psychological state and then measured the consequences. There's a wish now to have everything be automated so it can be done quickly and easily online." These days, he continues, there's less and less actual behavior in the science of behavior. "It's just sitting at a computer and doing readings."

I'm more inclined than Baumeister to see this replication failure as something truly momentous. Let's say it's true the tasks were wrong, and that ego depletion, as it's been described, is a real thing. If that's the case, then the study clearly shows that the effect is not

as sturdy as it seemed. One of the idea's major selling points is its flexibility: Ego depletion applied not just to experiments involving chocolate chip cookies and radishes, but to those involving word games, conversations between white people and black people, decisions on whether to purchase soap, and even the behavior of dogs. In fact, the incredible range of the effect has often been cited in its favor. How could so many studies, performed in so many different ways, have all been wrong?

Yet now we know that ego depletion might be very fragile. It might be so sensitive to how a test is run that switching from a pen and paper to a keyboard and screen would be enough to make it disappear. If that's the case, then why should we trust all those other variations on the theme? If that's the case, then the Big Idea has shrunk to something very small.

The diminution of the Big Idea **isn't easy to accept**, even for those willing to concede that there are major problems in their field. An ego depletion optimist might acknowledge that psychology studies tend to be too small to demonstrate a real effect, or that scientists like to **futz around with their statistics** until the answers come out right. (None of this implies deliberate fraud; just that sloppy standards prevail.) Still, the optimist would say, it seems unlikely that such mistakes would propagate so thoroughly throughout a single literature, and that so many noisy, spurious results could line up quite so perfectly. If all these successes came about by random chance, then it's a miracle that they're so consistent.

And here's the pessimist's counterargument: It's easy to imagine how one bad result could lead directly to another. Ego depletion is such a bold, pervasive theory that you can test it in a thousand different ways. Instead of baking up a tray of chocolate chip cookies, you can tempt your students with **an overflowing bowl of M&Ms**. Instead of having subjects talk to people of another race, you can ask them to **recall a time that they were victimized by racism**. Different versions of the standard paradigm all produce the same effect—that's the nature of the Big Idea. That means you can tweak the concept however you want, and however many times you need, until you've stumbled on a version that seems to give a positive result. But then your replication of the concept won't always mean you have a real result. It will only show that you've tried a lot of different methods—that you had the willpower to stick with your hypothesis until you found an experiment that worked.

Taken at face value, the new Registered Replication Report doesn't invalidate everything we thought we knew about willpower. A person's self-control can lapse, of course. We just don't know exactly when or why. It might even be the case that Baumeister has it exactly right—



that people hold a reservoir of mental strength that drains each time we use it. But the two-task method that he and Tice invented 20 years ago now appears to be in doubt. As a result, an entire literature has been rendered suspect.

“At some point we have to start over and say, *This is Year One*,” says Inzlicht, referring not just to the sum total of ego depletion research, but to how he sometimes feels about the entire field of social psychology.\*

All the old methods are in doubt. Even meta-analyses, which once were thought to yield a gold standard for evaluating bodies of research now seem somewhat worthless. “Meta-analyses are fucked,” Inzlicht warned me. If you analyze 200 lousy studies, you’ll get a lousy answer in the end. It’s garbage in, garbage out.

Baumeister, for his part, intends to launch his own replication effort, using methods that he thinks will work. “We try to do straight, honest work, and now we have to go to square one—just to make a point that was made 20 years ago. ... It’s easier to publish stuff that tears something down than it is to build something up,” he told me wearily. “It’s not an enjoyable time. It’s not much fun.”

If it’s not much fun for the people whose life’s work has been called into question, neither does it hearten skeptics in the field. “**I’m in a dark place**,” Inzlicht wrote on his blog earlier this week. “I feel like the ground is moving from underneath me and I no longer know what is real and what is not.”

**\*Correction, March 7, 2016:** This piece originally misquoted Michael Inzlicht as saying, in reference to the field of social psychology, “At this point we have to start over.” He said, “At some point we have to start over.” (**Return.**)

**PROMOTED STORIES**



**Fast Food You Should Never Order According to Staff**  
Interstice



**8 Things You Probably Didn't Know About CrossFit**  
Interstice



**10 Awesome Technologies That Exist Now**  
Tegzy



**6,000-year-old 'wand' discovered in cave**  
From the Grapevine

SCIENCE THE STATE OF THE UNIVERSE.

JULY 25 2016 9:39 AM

# Was This Study Even Peer-Reviewed?

How journalists should write about unpublished scientific research.

By Daniel Engber



Protesters struggle with police in Cleveland on Wednesday.

Spencer Platt/Getty Images

Amid the nightmare of the past three weeks, with black men being shot by cops, and then black men shooting cops, the *New York Times* ran a surprising front-page headline: “**Analysis Finds No Racial Bias in Lethal Force.**” Nestled between articles on killings in Baton Rouge, Louisiana; Dallas; and Staten Island, New York, the story described research by Harvard economist Roland Fryer, whose team scoured reams of data on shootings by police in Houston and concluded—among other things—that cops may be *less* likely to shoot at black suspects in any given situation. “**It is the most surprising result of my career,**” Fryer told the *Times*.



DANIEL ENGBER

Daniel Engber is a columnist for *Slate*.

---

Both Fryer’s study and the *Times*’ coverage of it were **widely criticized**. The newspaper never should have hyped the “**nonsensical Harvard study,**” tweeted *Washington Post* reporter Wesley Lowery, who won a Pulitzer Prize earlier this year for his reporting on killings by police. He added: “**Since when are economists authorities on police shootings?**” The venerable debunking website Snopes.com weighed in to discredit Fryer’s paper, even putting **scare quotes** around its description of the “Harvard study.” (Those words lent the work “a gravitas that was not yet present,” the site explained.) Complaints from readers who were “**confused, angry or both**” also reached the desk of the *New York Times*’ public editor. “The authors could have worked harder to anticipate confusion,” Liz Spayd wrote.

**CONTINUE READING**

FOLLOW SLATE

SLATE ON IPAD IPHONE ANDROID KINDLE

REPRINTS

ADVERTISE WITH US

ABOUT US  
CONTACT US

WORK WITH US  
USER AGREEMENT

[PRIVACY POLICY](#)

[FAQ](#)

[FEEDBACK](#)

[CORRECTIONS](#)

Slate is published by The Slate Group, a Graham Holdings Company. All contents © 2016 The Slate Group LLC. All rights reserved.