

Making It All Up

The behavioral sciences scandal

Andrew Ferguson

October 19, 2015, Vol. 21, No. 06

One morning in August, the social science reporter for National Public Radio, a man named Shankar Vedantam, sounded a little shellshocked. You couldn't blame him.

Like so many science writers in the popular press, he is charged with reporting provocative findings from the world of behavioral science: “. . . and researchers were very surprised at what they found. The peer-reviewed study suggests that [dog lovers, redheads, Tea Party members] are much more likely to [wear short sleeves, participate in hockey fights, play contract bridge] than cat lovers, but only if [the barometer is falling, they are slapped lightly upside the head, a picture of Jerry Lewis suddenly appears in their cubicle . . .].”

I'm just making these up, obviously, but as we shall see, there's a lot of that going around.

On this August morning *Science* magazine had published a scandalous article. The subject was the practice of behavioral psychology. Behavioral psychology is a wellspring of modern journalism. It is the source for most of those thrilling studies that keep reporters like Vedantam in business.

Over 270 researchers, working as the Reproducibility Project, had gathered 100 studies from three of the most prestigious journals in the field of social psychology. Then they set about to redo the experiments and see if they could get the same results. Mostly they used the materials and methods the original researchers had used. Direct replications are seldom attempted in the social sciences, even though the ability to repeat an experiment and get the same findings is supposed to be a cornerstone of scientific knowledge. It's the way to separate real information from flukes and anomalies.

These 100 studies had cleared the highest hurdles that social science puts up. They had been edited, revised, reviewed by panels of peers, revised again, published, widely read, and taken by other social scientists as the starting point for further experiments. Except . . .

The researchers, Vedantam glumly told his NPR audience, “found something very disappointing. Nearly two-thirds of the experiments did not replicate, meaning that scientists repeated these studies but could not obtain the results that were found by the original research team.”

“Disappointing” is Vedantam's word, and it was commonly heard that morning and over the following several days, as the full impact of the project's findings began to register in the world of social science. Describing the Reproducibility Project's report, other social psychologists, bloggers, and science writers tried out “alarming,” “shocking,” “devastating,” and “depressing.”

But in the end most of them rallied. They settled for just “surprised.” Everybody was surprised that two out of three experiments in behavioral psychology have a fair chance of being worthless.

The most surprising thing about the Reproducibility Project, however—the most alarming, shocking, devastating, and depressing thing—is that anybody at all was surprised. The warning bells about the feebleness of behavioral science have been clanging for many years.

THINK POSITIVE

For one thing, the “reproducibility crisis” is not unique to the social sciences, and it shouldn't be a surprise it would touch social psychology too. The widespread failure to replicate findings has afflicted physics, chemistry, geology, and other real sciences. Ten years ago a Stanford researcher named John Ioannidis published a paper called “Why Most Published Research Findings Are False.”

“For most study designs and settings,” Ioannidis wrote, “it is more likely for a research claim to be false than true.” He used medical research as an example, and since then most systematic efforts at replication in his field have borne him out. His main criticism involved the misuse of statistics: He pointed out that almost any pile of data, if sifted carefully, could be manipulated to show a result that is “statistically significant.”

Statistical significance is the holy grail of social science research, the sign that an effect in an experiment is real and not an accident. It has its uses. It is indispensable in opinion polling, where a randomly selected sample of people can be statistically enhanced and then assumed to represent a much larger population.

But the participants in behavioral science experiments are almost never randomly selected, and the samples are often quite small. Even the wizardry of statistical significance cannot show them to be representative of any people other than themselves.

This is a crippling defect for experiments that are supposed to help predict the behavior of people in general. Two economists recently wrote a little book called *The Cult of Statistical Significance*, which demonstrated how easily a range of methodological flaws can be obscured when a researcher strains to make his experimental data statistically significant. The book was widely read and promptly ignored, perhaps because its theme, if incorporated into behavioral science, would lay waste to vast stretches of the literature.

Behavioral science shares other weaknesses with every field of experimental science, especially in what the trade calls “publication bias.” A researcher runs a gauntlet of perverse incentives that encourages him to produce positive rather than negative results. Publish or perish is a pitiless mandate. Editors want to publish articles that will get their publications noticed, and researchers, hoping to get published and hired, oblige the tastes of editors, who are especially pleased to gain the attention of journalists, who hunger for something interesting to write about.

Negative results, which show that an experiment does not produce a predicted result, are just as valuable scientifically but unlikely to rouse the interest of Shankar Vedantam and his colleagues. And positive results can be got relatively easily. Behavioral science experiments yield mounds of data. A researcher assumes, like the boy in the old joke, that there must be a pony in there somewhere. After some data are selected and others left aside, the result is often a “false positive”—interesting if true, but not true.

Publication bias, compounded with statistical weakness, makes a floodtide of false positives. “Much of the scientific literature, perhaps half, may simply be untrue,” wrote the editor of the medical journal *Lancet* not long ago. Following the Reproducibility Project, we now know his guess was probably too low, at least in the behavioral sciences. The literature, continued the editor, is “afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance.”

Behavioral science suffers from these afflictions only more so. Surveys have shown that published studies in social psychology are five times more likely to show positive results—to confirm the experimenters’ hypothesis—than studies in the real sciences.

This raises two possibilities. Either behavioral psychologists are the smartest researchers, and certainly the luckiest, in the history of science—or something is very wrong. And we don’t have to assume bad faith on the part of social scientists, although it helps. The last three years have brought several well-publicized cases of prominent researchers simply making up data. An anonymous poll four years ago showed that 15 percent of social psychologists admitted using questionable research practices, from overmassaging their data to fabricating it outright. Thirty percent reported they had seen firsthand other researchers do the same.

THE LAB KIDS

Behavioral science has many weaknesses unique to itself. Remember that the point of the discipline is to discover general truths that will be useful in predicting human behavior. More than 70 percent of the world’s published psychology studies are generated in the United States. Two-thirds of them draw their subjects exclusively from the pool of U.S. undergraduates, according to a survey by a Canadian economist named Joseph Henrich and two colleagues. And most of those are students who enroll in psychology classes. White, most of them; middle- or upper-class; college educated, with a taste for social science: not John Q. Public.

This is a problem—again, widely understood, rarely admitted. College kids are irresistible to the social scientist: They come cheap, and hundreds of them are lying around the quad with nothing better to do. Taken together, Henrich and his researchers said, college students in the United States make “one of the worst subpopulations one could study for generalizing about *Homo sapiens*.” Different nationalities show a large variation in precisely the kinds of responses today’s social scientists love to study and generalize about: sexual and racial biases, expectations about the effects of power and money, attitudes toward social cooperation, habits of moral reasoning—even “spatial cognition.” Deep differences are found as well within subpopulations of Americans according to age, income, geographical upbringing, and educational levels.

“U.S. undergraduates exhibit demonstrable differences, not only from non-university educated Americans, but even from previous generations of their own families,” Greg Downey wrote of Henrich’s study. The variations are found even between groups of college students, and from one college to the next. Henrich wryly suggested a prominent journal change its name to *Journal of Personality and Social Psychology of American Undergraduate Psychology Students*.

“Despite their narrow samples,” Henrich wrote, “behavioral scientists often are interested in drawing inferences about the *human* mind. This

inferential step is rarely challenged or defended—with some important exceptions—despite the lack of any general effort to assess how well results from [their] samples generalize to the species. This lack of epistemic vigilance underscores the prevalent, though implicit, assumption that the findings one derives from a particular sample will generalize broadly.”

The defenders of behavioral science like to say it is the study of “real people in real-life situations.” In fact, for the most part, it is the study of American college kids sitting in psych labs. And the participation of such subjects is complicated from the start: The undergrads agree to become experiment fodder because they are paid to do so or because they’re rewarded with course credit. Either way, they do what they do for personal gain of some kind, injecting a set of motivations into the lab that make generalizing even riskier.

THE DANGERS OF MONOCHROME

Behind the people being experimented upon are the people doing the experimenting, the behavioral scientists themselves. In important ways they are remarkably monochromatic. We don’t need to belabor the point. In a survey of the membership of the Society for Personality and Social Psychology, 85 percent of respondents called themselves liberal, 6 percent conservative, 9 percent moderate. Two percent of graduate students and postdocs called themselves conservative. “The field is shifting leftward,” wrote one team of social psychologists (identifying themselves as “one liberal, one centrist, two libertarians, two who reject characterization,” and no conservatives). “And there are hardly any conservative students in the pipeline.” A more recent survey of over 300 members of another group of experimental psychologists found 4 who voted for Mitt Romney.

The self-correction essential to science is less likely to happen among people whose political and cultural views are so uniform. This is especially true when so many of them specialize in studying political and cultural behavior. Their biases are likely to be invisible to themselves and their colleagues. Consider this abstract from a famous study on conservatism [with technical decoration excised]:

A meta-analysis confirms that several psychological variables predict political conservatism: death anxiety; system instability; dogmatism—intolerance of ambiguity; openness to experience; uncertainty tolerance; needs for order, structure, and closure; integrative complexity; fear of threat and loss; and self-esteem. The core ideology of conservatism stresses resistance to change and justification of inequality and is motivated by needs that vary situationally and dispositionally to manage uncertainty and threat.

Only a scientist planted deep in ideology could read such a summary and miss the self-parodic assumptions buried there. Yet few people in behavioral sciences bat an eye. “Political Conservatism as Motivated Social Cognition,” which this paragraph is taken from, has been cited by nearly 2,000 other studies, accepted as a sober, scientific portrait of the “conservative” temperament.

In his book *Moral, Believing Animals*, Christian Smith, a sociologist at Notre Dame, described the worldview that undergirds politicized social science.

“Once upon a time,” Smith writes, describing the agreed-upon “narrative,” “the vast majority of human persons suffered in societies and social institutions that were unjust, unhealthy, repressive, and oppressive. These traditional societies were reprehensible because of their deep-rooted inequality, exploitation, and irrational traditionalism.”

Now, however, conditions have been improved, a bit, after much struggle. And yet the struggle goes on. “There is much work to be done to dismantle the powerful vestiges of inequality, exploitation, and repression.” Behavioral science, in this view, is part of the ongoing project of redress. It can counteract the psychological processes by which the powerful subjugate the powerless.

CONFIRMING CONFORMITY

Aping the forms and methods of physical scientists, crusading social scientists are bound to produce a lot of experiments that are quasi-scientific. They will resist replication if only because an experiment is just a one-off, a way to agitate and persuade rather than to discover. Scientists themselves speak of “confirmation bias,” an unnecessary term for a common human truth: We tend to believe what we want to believe.

When researchers, journal editors, peer-review panels, colleagues, and popular journalists share the same beliefs, confirmation bias will flourish. It’s human nature! Reading a typical behavioral science study involving race or sex, privilege or wealth or power, you can find it hard to distinguish between the experimenters’ premises and their conclusions. Better to scan the literature for what lawyers call “admissions against interest”—findings that contradict the prevailing creed. These are rare, but they exist, and they have undermined much of what behavioral scientists think they know about human behavior.

The subversion comes in many forms. A finding can often be undermined simply by looking closely at how the experiment was performed. Perhaps the most famous experiment in all of social science—I think we’re supposed to call it iconic—was undertaken in the early 1960s by Stanley Milgram, an assistant professor at Yale. Milgram was struck by the trial of the Nazi mass-murderer Adolf Eichmann, then underway in Israel.

His hunch was that Eichmann wasn’t singularly evil but merely a cog in the Nazi machine—a petty little man following great big orders. Nearly anyone, Milgram mused, could be induced to override his conscience and perform evil acts if he were instructed to do so by a sufficiently

powerful authority. Even someone from Yale.

This has since become known as conformity (not confirmation) bias, another elaborate and unnecessary verbalism invented to describe a home truth: We crave the approval of our friends and families, of people we take to be like ourselves. But conformity bias has been stretched much farther. The enormous power it holds to guide our behavior has crystallized as a settled fact in behavioral science.

To test his theory Milgram told his subjects that they were participating in a study of “learning.” A man in a lab coat took the subjects one at a time into a room and told them to turn an electric dial to shock a stranger in a room next door. They were to increase the strength of the shock by increments, finally to the point of inflicting severe pain. (The shock generator was a dummy; no one was actually hurt.)

The results were an instant sensation. The *New York Times* headline told the story: “Sixty-five Percent in Test Blindly Obey Order to Inflict Pain.” Two out of three of his subjects, Milgram reported, had cranked the dial all the way up when the lab-coat guy insisted they do so. Milgram explained the moral, or lack thereof: The “chief finding” of his study, he wrote, was “the extreme willingness of adults to go to almost any lengths on the command of an authority.” Milgram, his admirers believed, had unmasked the Nazi within us all.

Did he? A formidable sample of more than 600 subjects took part in his original study, Milgram said. As the psychologist Gina Perry pointed out in a devastating account, *Beyond the Shock Machine*, the number was misleading. The 65 percent figure came from a “baseline” experiment; the 600 were spread out across more than a dozen other experiments that were variations of the baseline. A large majority of the 600 did not increase the voltage to inflict severe pain. As for the participants in the baseline experiment who did inflict the worst shocks, they were 65 percent of a group of only 40 subjects. They were all male, most of them college students, who had been recruited through a newspaper advertisement and paid \$4.50 to participate.

The famous 65 percent thus comprised 26 men. How we get from the 26 Yalies in a New Haven psych lab to the antisemitic psychosis of Nazi Germany has never been explained.

Many replications of the Milgram experiments have succeeded, many have failed. But its importance to behavioral science cannot be overstated. It helped establish an idea that lies at the root of social psychology: Human beings are essentially mindless creatures at the mercy of internal impulses and outside influences of which they’re unaware. We may think we know what we’re doing most of the time, that we obey our consciences more often than not, that we can usually decide to do one thing and not another according to our own will. Behavioral scientists insist they know otherwise. This is the “mindlessness bias,” a just-invented (by me) term to describe the tendency of social psychologists to believe that their subjects are chumps.

There are other interpretations of Milgram’s results, after all, that are not quite so insulting to human nature. Perhaps the shockers were indeed conscious moral agents; maybe they had been persuaded they were participating in Science and, given the unlikelihood that a Yale Ph.D. student would let them cause harm, they were willing to do what they were told to advance the noble cause. Later interviews showed that most subjects thought this at the time of the experiments and were glad they had participated for precisely this reason. Others said they assumed the experiment was a ruse but went along anyway—some because they didn’t want to disappoint that nice man in the lab coat, some because they worried they might not get the \$4.50. The theory that they did what they did “blindly,” as the *Times* headline said, is an assumption, not a finding.

As one would-be replicator of the Milgram experiment, a heterodox researcher named Michael Shermer, wrote: “Contrary to Milgram’s conclusion that people blindly obey authorities to the point of committing evil deeds because we are so susceptible to environmental conditions, I saw in our subjects a great behavioral reluctance and moral disquietude every step of the way.”

PRIMED FOR SUCCESS

Milgram’s experiment has fared better than other demonstrations of the mindlessness bias. For a generation now, “priming” has been a favorite way for behavioral psychologists to demonstrate the slack-jawed credulity of their subjects. The existence of “perceptual priming” is well established among real scientists: If you talk to a person about zebras, he will be more likely to pick out the word “stripes” from a word jumble than a person who wasn’t thinking about zebras. This is an unconscious tendency that human beings undoubtedly possess, and the mechanism behind it is well understood by cognitive scientists.

But social psychologists set out to prove that social behavior, not just perception, could be determined by priming. Thanks to the pop science writer Malcolm Gladwell, who wrote about it in his bestselling book *Blink*, the most famous experiment in support of social priming involved a group of psych students (who else?) from New York University.

Thirty students were each given groups of words and told to arrange the words into coherent sentences. One set of students were given words that might be associated with old people: Florida, bingo, wrinkle. . . . The other group wasn’t primed; they were given “neutral” words, like thirsty, clean, and private. None of the students was told the true purpose of the experiment. (Behavioral science usually requires researchers to deceive people to prove how easily people can be deceived.)

As students finished the task, a researcher used a hidden stopwatch to measure how quickly each walked from the lab. The ones who hadn’t been primed took an average of 7.23 seconds to walk down the hallway. The ones who had been primed with the aging words took 8.28 seconds. The second group walked slower, just like old people! The kids couldn’t help themselves.

The researchers barely disguised their feelings of triumph.

“It remains widely assumed,” they wrote in their study, “that behavioral responses to the social environment are under conscious control.” But not anymore, not after this! Here science had discovered the “automatic behavior effect.” Some scientists call it “automaticity” for short. If you subtly set people up with the right cues, they’ll start doing things, without thinking, that they didn’t even know they were doing.

“The implications . . .,” wrote the researchers, “would appear to be formidable.”

Oh, they were. Behavioral science went priming crazy. Since its publication 20 years ago, “Automaticity of Social Behavior” has been cited in more than 3,700 published studies—an average of more than 15 studies a month, a staggering figure. Social psychologists discovered that priming empowered them to trick their subjects in many wonderful ways.

One experiment found that students primed with words about honesty became more honest; another found that if you fed them “achievement-related words” they would do better on achievement tests. And you didn’t need to use words. Another experiment found that subjects who held a heavy clipboard while conducting a job interview took social problems more seriously. If you showed subjects a picture of a college professor, their test scores improved. They reported feeling closer to their families if you showed them a graph with two points close together; if the points were farther apart, they said they felt more distant from their families.

The wonders rolled on. Students who were asked to type a morally questionable passage subsequently rated cleansing products more highly than other consumer products. (Because they felt dirty, see?) If they were required to wash their hands after describing an unethical act, they were less likely to offer help to someone who asked for it. (Because they had already washed away their guilt and felt no need to atone.) If they talked about an unethical act, they chose an offered bottle of mouthwash over hand sanitizer. (Because, without knowing it, they wanted to wash their mouths out.) And of course they preferred hand sanitizer over mouthwash if they wrote about it by hand. Because they unconsciously thought their hands were dirty.

Perhaps the summit of priming experiments was reached by researchers at Cornell. They contended that political conservatives were more obsessed with cleanliness than liberals. And sure enough: When they placed students closer to a bottle of cleanser, the students became more conservative in their views. Science is amazing.

TOO GOOD TO CHECK

It was left to Gladwell to summarize the sorry truth that priming revealed about ourselves. The experiments, he wrote, were “disturbing” because they “suggest that what we think of as free will is largely an illusion: much of the time, we are simply operating on automatic pilot, and the way we think and act—and *how well* we think and act on the spur of the moment—are a lot more susceptible to outside influences than we realize.”

Book buyers found the news of their own mindlessness more titillating than disturbing. *Blink* sold more than two million copies. Gladwell became a sage to the nation’s wealthiest and most powerful businessmen and policymakers, bringing them the latest word from the psych labs—a hipster version of the ancient soothsayers placing chicken entrails before the emperor.

Behavioral scientists hope that news of the automatic behavior effect will corrode the ordinary person’s self-understanding as a rational being in control of himself, more or less. Automaticity works all the way down, to where our moral views take shape. “Moral judgment . . .,” wrote one behavioral scientist recently, is “a kind of rapid automatic process more akin to the judgments animals make as they move through the world, feeling themselves drawn toward or away from various things.”

The question that Gladwell and his fellow journalists never asked, at least in public, was the crucial one: Are social-priming effects real? Do the results from the experiments truly constitute scientific knowledge? Or are they, as we journalists say, too good to check? The original researchers say their study has “been replicated dozens of times all over the world.”

This is true enough. But these studies are “conceptual replications”—experiments that are loosely modeled on the original experiment, or that simply accept its finding as a starting point for further manipulations.

Direct replications, on the other hand, including those for the Reproducibility Project, have failed to find a social-priming effect. Of seven replications undertaken for a project in 2014, only one succeeded. Again, no reason to be surprised. If “Automaticity of Social Behavior” hadn’t fit so snugly into the ideology of social science, the weaknesses of the original study would have been more readily seen.

For starters, the samples used in the study were small—the finding from one version of the experiment turned on the behavior of 13 students. If you want to make split-second measurements, you should be able to do better than to put a stopwatch in the hands of a furtive grad student. The opportunities for the researchers to signal the true purpose of the experiment to the subjects, or otherwise to influence the outcome, were too great to ignore. The reported effects were themselves weak, and they began to appear robust only after statistical enhancement. And no one, then or now, has been able to explain how social priming was supposed to work, technically. What possible mechanism, whether physical or psychological, would cause a word with many different connotations (Florida) to trigger the same stereotype (old people) and result in the same behavior (slow walking) in many different subjects?

When three researchers undertook a direct replication of the original social-priming experiment in 2012, they used a much larger sample in

hopes of getting stronger effects. To measure walking speeds, they replaced the stopwatch with infrared signals and automatic timers.

When their replication found no evidence for social priming, they did something devilish. They turned their attention to their own researchers, the people hired to test the subjects. They did another experiment. Half of their experimenters were led to believe that the primed subjects would walk slowly, half were told the primed subjects would walk quickly. The first group of experimenters were far more likely to find a slow-walking effect than the second group of experimenters.

If there was a priming effect, in other words, it was operating on the people doing the experiment, not on the people being experimented upon. An earlier high-profile experiment, again using sensors, automatic timers, and a larger sample, likewise failed to confirm the slow-walking effect.

The lead author of “Automaticity,” John Bargh of Yale, responded as a dogmatist would, with ad hominem attacks on the researchers, accusing them of bad faith, incompetence, and a terrible case of (dis)confirmation bias. He and his defenders pointed to the many conceptual replications and argued, correctly, that a failure to reproduce an effect is not the same as proving it doesn’t exist.

But the bad news for social priming keeps coming. The lead researcher on one of the social-priming replications next turned to “social distance” priming—the experiments judging familial closeness with points on a graph. It was a direct replication, and it failed. Next came a direct replication of the experiment showing that priming with “achievement words” led to higher achievement. No such result could be reproduced. An attempt to replicate the “honesty” experiment came next. Does speaking words about honesty make you more honest? The original finding had become a staple of the literature, cited in 1,100 studies. No effect was found in the direct replication.

The researchers concluded drily: “These failures to replicate, along with other recent results, suggest that the literature on goal priming requires some skeptical scrutiny.” Because of the tendency to publish only positive results and ignore negative findings, “a published literature can easily provide a very misleading picture of reality.” And if many kinds of priming can’t be replicated under supposedly controlled laboratory conditions, how predictable could the effect be in the kaleidoscope of daily life, where human beings are battered and pummeled by an infinite number of influences? Malcolm Gladwell, meanwhile, has moved on to other things.

EPIDEMIC OF FAILURE

Even before the Reproducibility Project, direct replications failed to find evidence for many other effects that the social psychology literature treats as settled science. “Single-exposure conditioning”—if you’re offered a pen while your favorite music is playing, you’ll like the pen better than one offered while less appealing music plays. The “primacy of warmth effect,” which tells us our perceptions are more favorable to people described as “warm” than to people described as “competent.” The “Romeo and Juliet effect”: Intervention by parents in a child’s romantic relationship only intensifies the feelings of romance. None of these could be directly replicated.

Perhaps most consequentially, replications failed to validate many uses of the Implicit Association Test, which is the most popular research tool in social psychology. Its designers say the test detects unconscious biases, including racial biases, that persistently drive human behavior. Sifting data from the IAT, social scientists tell us that at least 75 percent of white Americans are racist, whether they know it or not, even when they publicly disavow racial bigotry. This implicit racism induces racist behavior as surely as explicit racism. The paper introducing the IAT’s application to racial attitudes has been cited in more than 6,600 studies, according to Google Scholar. The test is commonly used in courts and classrooms across the country.

That the United States is in the grip of an epidemic of implicit racism is simply taken for granted by social psychologists—another settled fact too good to check. Few of them have ever returned to the original data. Those who have done so have discovered that the direct evidence linking IAT results to specific behavior is in fact negligible, with small samples and weak effects that have seldom if ever been replicated. One team of researchers went through the IAT data on racial attitudes and behavior and concluded there wasn’t much evidence either way.

“The broad picture that emerges from our reanalysis,” they wrote, “is that the published results [confirming the IAT and racism] are likely to be conditional and fragile and do not permit broad conclusions about the prevalence of discriminatory tendencies in American society.” Their debunking paper, “Strong Claims and Weak Evidence,” has been cited in fewer than 100 studies.

MOUNTING A DEFENSE

Amid the rubble of the replication crisis, the faithful of social science have mounted a number of defenses. “Critics will be delighted to crow over the findings of low levels of reproducibility,” one wrote in the *Guardian*. “However, the crowing might die down when it is pointed out that problems of reproducibility have been raising alarm bells in many other areas of science, including some much ‘harder’ subjects.”

Defenders cited a host of biases to which the original researchers might have succumbed, especially publication bias and selective data bias. And besides, the defenders pointed out, a failed replication doesn’t tell us too much: The original study might be wrong, the replication might be wrong, they might both be wrong or right. Small changes in methodology might influence the results; so might the pool of people the samples are drawn from, depending on age, nation of origin, education level, and a long list of other factors. The original study might be reproducible in certain environments and not in others. The skill of the individual researcher might enter in as well.

All true! Rarely do social scientists concede so much about the limitations of their trade; the humility is as welcome as it is unexpected. But

these are not so much defenses of social psychology as explanations for why it isn't really science. If the point is to discover universal tendencies that help us predict how human beings will behave, then the fragility of its experimental findings renders them nearly useless. The chasm that separates the psych lab from everyday life looks unbridgeable. And the premise of behavioral science—that the rest of us are victims of unconscious forces that only social scientists can detect—looks to be not merely absurd but pernicious.

For even as it endows social scientists with bogus authority—making them the go-to guys for marketers, ideologues, policymakers, and anyone else who strives to manipulate the public—it dehumanizes the rest of us. The historian and humanist Jacques Barzun noticed this problem 50 years ago in his great book *Science: The Glorious Entertainment*. Social psychology proceeds by assuming that the objects (a revealing word) of its study lack the capacity to know and explain themselves accurately. This is the capacity that makes us uniquely human and makes self-government plausible. We should know enough to be wary of any enterprise built on its repudiation.

This is probably why humility among social scientists never lasts; it's not in the job description. No sooner do social scientists concede the limitations of their work than they begin the exaggerations again. A week after the Reproducibility Project set off its cluster bomb, President Obama's Social and Behavioral Sciences Team issued its first annual report. (Who knew there was such a thing?) The team describes itself as "a cross-agency group of experts in applied behavioral science that translates findings and methods from the social and behavioral sciences into improvements in Federal policies and programs."

We can be relieved that the work of the team is much less consequential than it sounds. So far, according to the report, the team has made two big discoveries. First, reminding veterans, via email, about the benefits they're entitled to increases the number of veterans applying for the benefits. Second, if you simplify complicated application forms for government financial aid—for college students and farmers, let's say—the number of students and farmers who apply for financial aid will increase. "One behaviorally designed letter variant" increased the number of farmers asking for a microloan "from 0.09 to 0.11 percent."

Evidently impressed with all this science, President Obama issued an executive order directing federal agencies "to use behavioral insights to better serve the American people." Agency heads and personnel directors were instructed to "recruit behavioral science experts to join the Federal government as necessary to achieve the goals of this directive." We should have known! After all the bogus claims and hyped findings and preening researchers, after the tortured data and dazed psych students, this is the final product of the mammoth efforts of behavioral science: a federal jobs program for behavioral scientists.

ONWARD

A few days after his report on the Reproducibility Project, Shankar Vedantam was back at his post. He sounded much better, and with good reason: He had found a new study. Israeli researchers had examined why girls, who do better than boys on math tests, shy away from math courses when they get to high school. This is a very hot topic in social science, and in journalism.

Perhaps chastened by the findings of the Reproducibility Project, Vedantam told listeners that the study would someday need to be replicated, but for now . . .

"The new study suggests," Vedantam said, "that some of these outcomes might be driven by the unconscious bias of elementary school teachers."

"Suggests . . . some of . . . might be . . ." He was showing admirable restraint.

But then he must have figured, what the hell. In the rest of his report he treated the bias as unassailable fact.

So did the researchers. NPR listeners, if they had the energy, could have downloaded the study for themselves. They would have seen firsthand that the study compared apples and oranges, that it was statistically suspect, and that it recorded no instance of actual bias but simply assumed what it hoped to prove: that the bias of elementary school teachers was keeping women out of mathematics.

Vedantam showed great sympathy for these deluded teachers, most of them female, who were victimizing their female students without knowing it.

"It's hard to imagine that these teachers actually have conscious animosity toward the girls in their classroom," said NPR's social science reporter. "Much more likely these biases are operating at an unconscious level."

Not anymore! The headline over Vedantam's NPR blog said it all: "Hard Evidence: Teachers' Unconscious Biases Contribute to Gender Disparity."

Hand in hand with journalism, Science marches on and on.

Andrew Ferguson is a senior editor at The Weekly Standard.

[Subscribe now to The Weekly Standard!](#)

Get more from The Weekly Standard: [Follow WeeklyStandard.com on RSS](#) and [sign-up for our free Newsletter](#).

Copyright 2015 Weekly Standard LLC.

Source URL: http://www.weeklystandard.com/articles/making-it-all_1042807.html