

THE CHRONICLE

of Higher Education

Research

[Home](#) [News](#) [Faculty](#) [Research](#)



November 13, 2011

Fraud Scandal Fuels Debate Over Practices of Social Psychology

Even legitimate researchers cut corners, some admit

By Christopher Shea

The discovery that the Dutch researcher Diederik A. Stapel made up the data for dozens of research papers has shaken up the field of social psychology, fueling a discussion not just about outright fraud, but also about subtler ways of misusing research data. Such misuse can happen even unintentionally, as researchers try to make a splash with their peers—and a splash, maybe, with the news media, too.

Mr. Stapel's conduct certainly makes him an outlier, but there's no doubt he was a talented mainstream player of one part of the academic-psychology game: The now-suspended professor at Tilburg University, in the Netherlands, served up a diet of snappy, contrarian results that reporters lapped up.

Consider just two of his most recent papers: "[Power Increases Infidelity Among Men and Women](#)," from *Psychological Science*, and "[Coping With Chaos: How Disordered Contexts Promote Stereotyping and Discrimination](#)," from *Science*—two prestigious journals. The first paper upended a gender stereotype (alpha-female politicians philander, too?!), while the second linked the physical world to the psychological one in a striking manner (a

messy desk leads to racist thoughts!?). Both received extensive news coverage.

Even before the Stapel case broke, a flurry of articles had begun appearing this fall that pointed to supposed systemic flaws in the way psychologists handle data. But one methodological expert, Eric-Jan Wagenmakers, of the University of Amsterdam, added a sociological twist to the statistical debate: Psychology, he argued in a recent [blog post](#) and an interview, has become addicted to surprising, counterintuitive findings that catch the news media's eye, and that trend is warping the field.

"If high-impact journals want this kind of surprising finding, then there is pressure on researchers to come up with this stuff," says Mr. Wagenmakers, an associate professor in the psychology department's methodology unit.

Bad things happen when researchers feel under pressure, he adds—and it doesn't have to be Stapel-bad: "There's a slippery slope between making up your data and torturing your data."

In September, in comments [quoted](#) by the statistician Andrew Gelman on his blog, Mr. Wagenmakers wrote: "The field of social psychology has become very competitive, and high-impact publications are only possible for results that are really surprising. Unfortunately, most surprising hypotheses are wrong. That is, unless you test them against data you've created yourself."

Is a desire to get picked up by the Freakonomics blog, or the dozens of similar outlets for funky findings, really driving work in psychology labs? Alternatively—though not really mutually

exclusively—are there broader statistical problems with the field that let snazzy but questionable findings slip through?

Statistical Significance

Discovering important results in small samples of test subjects is always a tricky business, and psychologists who want to reform the field's practices have noted that much hinges on the statistical tools used.

To show just how easy it is to get a nonsensical but "statistically significant" result, three scholars, in an article in November's *Psychological Science* titled "[False-Positive Psychology](#)," first showed that listening to a children's song made test subjects feel older. Nothing too controversial there.

Then they "demonstrated" that listening to the Beatles' "When I'm 64" made the test subjects literally younger, relative to when they listened to a control song. Crucially, the study followed all the rules for reporting on an experimental study. What the researchers omitted, as they went on to explain in the rest of the paper, was just how many variables they poked and prodded before sheer chance threw up a headline-making result—a clearly false headline-making result.

The odds of statistical bogosity grow when researchers don't have to report all the ways they manipulated their data in exploratory fashion. For example, the researchers "used father's age to control for baseline age across participants," thereby fudging the subjects' actual ages. They factored in lots of completely irrelevant data. And, rather than establish from the outset how many subjects they would test, they tested until they obtained the false result.

The authors of that provocative paper were Joseph P. Simmons and Uri Simonsohn of the University of Pennsylvania, and Leif D. Nelson of the University of California at Berkeley. "Many of us," they wrote—"and this includes the three authors of this article"—end up "yielding to the pressure to do whatever is justifiable to compile a set of studies that we can publish. This is driven not by a willingness to deceive but by the self-serving interpretation of ambiguity. ... "

In a forthcoming paper, also to appear in *Psychological Science*, Leslie K. John, an assistant professor at Harvard Business School, and two co-authors report that about a third of the 2,000 academic psychologists they surveyed admit to questionable research practices. Those don't include outright fraud, but rather such practices as stopping the collection of data when a desired result is found, or omitting from the final paper some of the variables tested.

And Mr. Wagenmakers himself was an author of a paper this year, "Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi." It appeared in the *Journal of Personality and Social Psychology*, inspired by that journal's publication of a much-discussed, and much-ridiculed, paper on "psi," or psychic phenomena, like "precognition," or perceiving an event before it occurs.

The Cornell University psychologist Daryl Bem had reported evidence that people could predict the future at a better-than-chance rate under some circumstances—whether an image would appear on the left or right side of a screen, for instance. That such a hypothesis could be "proved" in labs, even though clearly no one is getting rich by deploying psi in casinos, was more than a little

problematic, Mr. Wagenmakers argued. Only dubious statistics could explain such a finding, he said.

The technical complaints about current statistical testing in psychology are by now familiar to those in the field. The standard measure of "statistical significance" is the "P value," which indicates the likelihood that a result is due to chance. By definition, a P value of 0.05 means there's a 1-in-20 likelihood the finding is a fluke. Add the researcher's freedom to explore multiple variables without reporting the extent of the searching in the final paper, and problems multiply. Add the so-called file-drawer effect—failed attempts to establish correlations seldom get published, but the odd lucky strike will—and the problems multiply further.

The Great Headline

Mr. Wagenmakers adds an argument involving a feedback loop between researchers looking for surprising findings and news media hungry to report them.

Unlike most other critics, he's not afraid to call out specific papers that he thinks are bogus: "Through prestigious publications and extensive media coverage," he writes in a draft of a new paper, a portion of which he shared with *The Chronicle*, "the general public has been informed that engineers have more sons and nurses have more daughters, ... that people choose spouses, places to live, and professions because they share letters with their name (e.g., Jenny marries Jim, Phil moves to Philadelphia, and Dennis becomes a dentist, ... that people make better decisions when their bladder is full, ... that ovulation makes it easier for women to distinguish heterosexual from homosexual men, ... and that brief exposure to an image of the American flag can push people toward the Republican

end of the U.S. political spectrum, even when the flag image was presented eight months earlier."

He can't swear all those studies are wrong. "But even using common sense, a lot of these hypotheses are unlikely, a priori, and you should collect a lot more evidence in order for them to be accepted."

Needless to say, the authors of the studies he alludes to demur. "I am insulted," writes Mirjam A. Tuk, author of "[Inhibitory Spillover: Increased Urination Urgency Facilitates Impulse Control in Unrelated Domains](#)," in an e-mail. The paper was published this year in *Psychological Science*. The idea that self-control in one area might contribute to self-control in a different arena is one rooted in neurological theory, explains Ms. Tuk, of the University of Twente, in the Netherlands. "Conducting serious, theoretically sound research is my primary aim, and by no means one I would ever trade off [for] press attention."

Travis Carter, a postdoctoral fellow at the Center for Decision Research at the University of Chicago's Booth School of Business, co-wrote the article on how [exposure to the American flag](#) affects voting behavior, which also appeared in *Psychological Science* this year. He says his team has done several studies that confirm the effects of flag exposure on political views, some of which may yet be published elsewhere, and adds, "We don't have a big file drawer full of failed studies."

Yet, interestingly, he does not reject Mr. Wagenmakers's broader argument: "I absolutely agree that people strive for the kind of studies that get media attention." Those studies are problematic, he says, in part because they often don't grow out of a broader theory, but rather amount to little more than, "Here's a quick little effect

that we can show." Studies like that "are more likely to be flukes," he says.

"I want to publish very high-quality work," he says, "but there's certainly a push to get more stuff out there. The temptations to cut corners are certainly there."

Eliot R. Smith, new editor of the *Journal of Personality and Social Psychology*, says the talk about psychologists pursuing "sexy" findings is way overblown. "Go through five issues of mainstream psychological journals," says Mr. Smith, a social psychologist at Indiana University at Bloomington. "You'll see maybe five articles out of 50 that are big counterintuitive findings that your grandmother would be interested in."

For most of the others, no one outside the relevant subfield would even understand the point of the experiment, let alone say "wow" at the result. He also doesn't see why someone interested in cutting corners would be any more likely to do so on a colorful topic than a "dull" one, of interest only to specialists. A publication is a publication, after all.

Robert V. Kail, editor of *Psychological Science*, says he's never heard of the likelihood of press attention being used as a reason to publish a researcher's work. Rather, he says, he asks his reviewers: "If you are a psychologist in a specialty area, is this the kind of result that is so stimulating or controversial or thought-provoking that you'd want to run down the hall and tell your colleagues in another subfield, 'This is what people in my field are doing, and it's really cool.'?"

"To me that's not 'sexy.' It's the most interesting science that we're doing," says Mr. Smith. And it might have to do with reaction times

or perception, not anything you'd read about in *The Wall Street Journal* or *The New York Times*. Moreover, the eye-catching studies may well be rooted in sound psychological theory—which Mr. Wagenmakers fails to mention in his drive-by attacks on specific papers, Mr. Smith says.

Research Reform

Since the extent of Mr. Stapel's misdeeds is not yet clear, it's too early to say what, if any, steps might be put in place to prevent future occurrences.

Still, reforms are in the works. Mr. Wagenmakers advocates an alternative to P-value testing, called Bayesian statistics, which incorporates such information as prior expectations that a hypothesis is true. (It's complex, but the bar for accepting something like psi would be higher, for starters.) That approach has some supporters, but it's not universally accepted, and it would require retraining both graduate students and the professors who teach them.

Mr. Simmons, Mr. Nelson, and Mr. Simonsohn, of the "When I'm 64" paper, recently met with the new editor of *Psychological Science*, Eric Eich, of the University of British Columbia, to push for some of the reforms they advocated for in their paper—namely, fuller descriptions of research protocols, and more tolerance of imperfections in initial papers. When the data are supposed to support a thesis perfectly, the incentives to cut corners increase.

Mr. Smith, the *Journal of Personality and Social Psychology* editor, describes such reforms as a natural part of any science. "There are problems with the way the field of psychology approaches statistical analysis," he says, "but my impression is there

is not a clear consensus that the whole field is doing it wrong and we should change."

And it should be said that other fields are convulsed with similar internal criticisms. For example, John Ioannidis, an epidemiologist at Stanford Medical School, has suggested that most medical studies are statistically flawed.

Mr. Wagenmakers says reform needs to happen more quickly. "The field is slowly being polluted by these errors," he says of the false positives. And social psychology is in danger of becoming risible. The article on urination and self-control, published in the flagship journal of the Association for Psychological Science, won an Ig Nobel Prize this year, a tongue-in-cheek recognition given by the magazine *Annals of Improbable Research* for achievements "that first make people laugh, and then make them think." But they tend to be bestowed on trivial-seeming work.


"If the work in key psychology journals starts to get these Ig Nobel prizes," he says, "it's something we have to worry about."

Comments
DISQUS
Powered by DISQUS



Add a comment

Log in to post
with your Chronicle account:

Don't have an account? [Create one now.](#)
Or log in using one of these alternatives:



Showing 17 comments

Sort by Follow comments:  by e-mail  by RSS

Real-time updating is **enabled.** ([Pause](#))



chrisyoung006 14 hours ago

Can they just do great work (great research) without needing affirmation? Most of the time, people believe in research that's why it's relatively important to be extra careful with the reports they were publishing.

Cheers!

[Rafael Apolinario III](#)

[Real Life Success Stories](#)

1 person liked this.

Like



nybound 13 hours ago

The root of the problem is that there is a 'people are so stupid' bias in much social scientific research (for a great article on this, see Kihlstrom (2004), Behavioral and Brain Sciences, 27, 348–348). An article that tells us that our brain works pretty well under most circumstances or that it is a pretty darn-good information processing machine just isn't 'interesting'. We have to hear about how we are all so stupid that that the downfall of society is imminent unless we understand this important piece of research!

By the time I finished my undergrad degree in psychology, I was amazed that the average person can get through the day without accidentally killing him/herself or someone else. Then I went to work in the 'real world' and met some incredibly smart people – like people who could eyeball a table of data produce virtually identical answers to my mine, which were based on hours of analyses and regressions. They weren't mathematical savants, they just learned to do this over time. When you think about it, we went from the first human flight in 1903 to being able to stage a remarkably convincing moon landing just 66 years later! ;)

In fact, our very success is our biggest threat right now, but we do seem to be waking up to that in time. The over-emphasis on biases and stupidity seems remarkably at odds with the notion that our cognitive heuristics should have evolved to be there precisely because they serve us so well. But we try desperately to concoct contrived experimental situations where these heuristics 'fail' (sometimes largely because the experimenter has misinterpreted what is in fact the 'rational' response) then bemoan how stupid the human being really is. At least, as elitist academics, it fits with our love of droning on and on about how stupid John Q. Public is, too stupid in fact to even to know what's best for himself.

5 people liked this.

Like



abd_punk 12 hours ago

C. Shea wrote "And, rather than establish from the outset how many subjects they would test, they tested until they obtained the false result."

What's wrong with sequential statistics? On average they minimize sample size, hence cost, for the desired significance level.

Like



catkelley 5 hours ago in reply to abd_punk

This practice violates an important assumption of statistical tests, that the observations are independent. It also capitalizes on chance findings. This is something you ought to learn in your very first elementary stastics course.

9 people liked this.

Like



antarcticchinstrap 2 hours ago in reply to abd_punk

Your thinking about stats may be a contributing factor to your ABD status as indicated by your name.

2 people liked this.

Like



blindboy 8 hours ago

There appears to be a wider re-evaluation of the analysis of data going on. The article mentions the field of epidemiology and I would suggest education as another field full of dubious results. The importance of getting the processes of analysis and publication right can hardly be over stated. It is not just the risk of damaging our understanding of the various fields. It undermines confidence in science more widely and encourages anti-intellectualism.

1 person liked this.

Like



Jeff Sherman 6 hours ago

It's bad enough to read this kind of poor reporting in the popular press, but truly sad to see it in an outlet specifically oriented toward academics. The two Psychological Science papers you cite that describe methodological trickery (the Simmons et al. & John papers) are not specific to social psychology, but describe problematic practices across areas of psychology. The title and tenor of this article is entirely misleading (what's that you wrote about Great Headlines?). If other fields of science cared to examine their practices as carefully as psychologists are now doing, undoubtedly, the results would be similar. Cold-fusion anybody? It's nice that, finally, at the very end of the article, you mention Ioannidis' work showing similar problems in other fields.

As for out-right fraud, there are data from the Office of Research

Integrity: <http://ori.hhs.gov/misconduct/...> These data indicate that psychology is far from a major

transgressor (biomedical research is where much of the action is).

Finally, I find your willingness to grant space for Wagenmakers to cast unsupported aspersions about specific findings to be deeply disappointing. I also would urge the reader to take Wagenmakers' comments about social psychology with a very large grain of salt. He is a cognitive psychologist whose comments clearly demonstrate that he is largely unfamiliar with the corpus of research in social psychology.

10 people liked this.



medendo 4 hours ago

Meta-analyses supposedly give a more accurate picture of research findings, but I have often complained that publication bias will always skew results. A study of outliers is no study at all. Maybe we need a Journal of Predictable Results to lend some balance to the field.

3 people liked this.



jdcarmine 3 hours ago

Hey? And what else is new in the world of the multi-billion dollar religion of badly controlled psych studies? How often do we need to see that the soft results of psychological sort-a-studies is a consequence of unreliable data pressed into the service of marginally quantitative studies for the sake of some socially preferred world view. Much of what passes for research in social psychology remains more faith-based than reason tested.

3 people liked this.



Jeff Sherman 20 minutes ago in reply to jdcarmine

Multi-billion dollar psych studies?! That's the single funniest thing I've read about this whole sordid affair.

Type your comment here.



Kate Ratliff 3 hours ago

Great response, Jeff.

I would like to note that Joris Lammers, the first author of the paper showing that powerful men and women are more likely than their less powerful counterparts to cheat, is standing behind that paper. He collected

those data himself and the Levelt committee investigating the Stapel incident has not named that paper as one that is suspected of fraud. The implication that that paper is false is irresponsible, unfair, and potentially damaging to the career of a young researcher who has done nothing wrong.

3 people liked this.

Like



jlmichaels 2 hours ago

Beyond the coy possibility that social psychologists are out for media glamor, there are at least two other sources for the statistics and data analysis woes that Shea fails to address. First, psychology training (and I'd suggest other fields relying on statistical analysis) widely lacks mathematical rigor. Second, inferential procedures reliant on data summary may not be capable of handling the complex and interactive systems at play in psychology.

For undergraduate and graduate training in psychology, it is typical to require only 2 to 4 "quantitative" courses for a given degree. I emphasize that "quantitative" can include courses with math that is so watered down that the students are trained to punch numbers into calculators and rely on rote reference to some formulae sheets without any understanding of why the formulae work, how they were derived, and what assumptions are at play with the equations.

The second potential source for data troubles in psychology is the fact that most statistics used in the field require that phenomena be translated from their complex reality to simple numerical summary. Compare this to research in physics or many earth and space sciences, where a given phenomena (e.g. a hurricane or galaxy) may be measured and modeled based on dozens if not hundreds of parameters and individual points. In psychology, it is common practice to translate the human mind and social interactions (interactions between minds) into single data points. This system of measurement in psychology is not necessarily the fault of the researchers. The lack of interaction between psychologists and cutting-edge mathematics may be one culprit. Another potential source for this trouble could be that journal editors and reviewers may be used to standard inferential approaches to data analysis and may be less willing to accept and publish research reliant on data analysis methods not often seen in traditional psychology work.

Finally, Shea's concerns should not be reserved for social psychology alone. For example, in cognitive psychology it is not uncommon to measure people's perception of stimuli in terms of how often the stimuli are correctly identified. This is often measured by "hit rate" (proportion of targets correctly identified). In such measurement, the "false alarm rate" (saying a stimulus is present when it is not) is often neglected, which begs the question of whether a high proportion of identified targets in a data set arises simply because a person makes more guesses or attempts.

While I disagree that social psychologists are out to get their name in newspapers or magazines I applaud Shea for bringing attention to statistics and research methodology, which are often seen as bland tools to be used rather than critical elements of the research process.

8 people liked this.

Like



davi2665 2 hours ago

Social psychology is by no means unique in having fabrication and fraudulent research. Many biomedical sciences areas have had similar problems. One of the more notorious involved anesthesiology/pain management, in which over 20 studies on multi-drug pain management were simply made up. Although the perpetrator is now in prison (appropriately), it still is not certain how many patients have experienced serious morbidity due to the lies and fabrications of the so-called researcher. Psychiatrists appear to be particularly involved in failing to report conflicts of interest- Charles Nemeroff is the poster child for that issue. What is needed is continued diligence to root out such misconduct, with IRBs and fellow investigators/reviewers in the lead. The response from NIH is at best weak and ineffective, with almost laughable "sanctions" for misconduct such as faking data (2-3 whole years without being able to apply for grants or sit on study sections). So until the NIH actually decides to take these matters seriously, it is up to individuals to pay attention and bring them to light.

2 people liked this. Like



justinmatus 1 hour ago

Reviewers are sometimes to blame as well. Not long ago an article I submitted was rejected for publication and I believe, in large part, because I used language that said the reader should interpret the results carefully, given the small sample size and a few other methodoligical concerns I voiced. The reviewer was essentially saying that I was undermining my study by pointing out the limitations of the study! I was trained that as a careful researcher you have an obligation to inform the reader to include the good, the bad and the ugly.

1 person liked this. Like



ddydek 1 hour ago

I saw a connection between this currently hot topic and the great assessment movement in higher education, which, in the service of accountability (or should I say self-service?), mandates that faculty members scrutinize data delivered from the Office of Institutional Effectiveness and come up with "actionable conclusions" based on those "findings." Some institutions even mandate that the findings WILL be "found." Is that intellectually honest? Or is it a set-up for fraud?

1 person liked this. Like



frankschmidt 29 minutes ago

I find it difficult to take a social psychology too seriously, given that so many of its conclusions are based on studies of college sophomores... ;)

Like



emmryss 0 minutes ago

Listening to "When I'm 64" made subjects "literally younger"? Now there's a result worthy of headlines!

Like

Copyright 2011. All rights reserved.

The Chronicle of Higher Education 1255 Twenty-Third St, N.W. Washington,