

What has happened down here is the winds have changed

Posted by [Andrew](#) on 21 September 2016, 9:03 am

Someone sent me [this article](#) by psychology professor Susan Fiske, scheduled to appear in the APS Observer, a magazine of the Association for Psychological Science. The article made me a little bit sad, and I was inclined to just keep my response [short and sweet](#), but then it seemed worth the trouble to give some context.

I'll first share the article with you, then give my take on what I see as the larger issues. The title and headings of this post allude to the fact that the [replication crisis has redrawn the topography of science](#), especially in social psychology, and I can see that [to people such as Fiske who'd adapted to the earlier lay of the land, these changes can feel catastrophic](#).

I will *not* be giving any sort of point-by-point refutation of Fiske's piece, because it's pretty much all about internal goings-on within the field of psychology (careers, tenure, smear tactics, people trying to protect their labs, public-speaking sponsors, career-stage vulnerability), and I don't know anything about this, as I'm an outsider to psychology and I've seen very little of this sort of thing in statistics or political science. (Sure, dirty deeds get done in all academic departments but in the fields with which I'm familiar, methods critiques are pretty much out in the open and the leading figures in these fields don't seem to have much problem with the idea that if you publish something, then others can feel free to criticize it.)

As I don't know enough about the academic politics of psychology to comment on most of what Fiske writes about, so what I'll mostly be talking about is how her attitudes, distasteful as I find them both in substance and in expression, can be understood in light of the recent history of psychology and its replication crisis.

Here's Fiske:

Mob Rule or Wisdom of Crowds?

Introduction by APS President Susan Goldin-Meadow: New forms of media are making it easier and easier for us to react to, and comment on, research within our community. Although free-flowing comments and criticisms can often push an argument or research program forward in a good direction, they can also derail, and perhaps even threaten, the process. I invited guest columnist Susan Fiske, a former APS president, to think about the impact that the new media are having not only on our science, but also on our scientists.

Susan T. Fiske

APS Past President

Our field has always encouraged – required, really – peer critiques. But the new media (e.g., blogs, twitter, Facebook posts) are encouraging uncurated, unfiltered trash-talk. In the most extreme examples, online vigilantes are attacking individuals, their research programs, and their careers. Self-appointed data police are volunteering critiques of such personal ferocity and relentless frequency that they resemble a denial-of-service attack that crashes a website by sheer volume of traffic.

Only what's crashing are people. These unmoderated attacks create collateral damage to targets' careers and well being, with no accountability for the bullies. Our colleagues at all career stages are leaving the field because of the sheer adversarial viciousness. I have heard from graduate students opting out of academia, assistant professors afraid to come up for tenure, mid-career people wondering how to protect their labs, and senior faculty retiring early, all because of methodological terrorism. I am not naming names because ad hominem smear tactics are already damaging our field. Instead, I am describing a dangerous minority trend that has an outsized impact and a chilling effect on scientific discourse. I am not a primary target, but my goal is to give voice to others too sensible to object publicly.

To be sure, constructive critics have a role, with their rebuttals and letters-to-the-editor subject to editorial oversight and peer review for tone, substance, legitimacy. Some moderated social-media groups monitor individual posts to ensure they are appropriate. Always, of course, if critics choose to write a personal message to the author, that's their business. If they request the original data, scientific norms demand delivery within reasonable constraints. All these venues respect the target.

What's more, APS has been a leader in encouraging robust methods: transparency, replication, power analysis, effect-size reporting, and data access. All this strengthens our field because APS innovates via expert consensus and explicit editorial policies. Individuals' research is judged through monitored channels, most

often in private with a chance to improve (peer review), or at least in moderated exchanges (curated comments and rebuttals). These venues offer continuing education, open discussion, and quality control. These constructive efforts draw on the volunteer talent of many, in the greater good and respecting the individual investigator.

In contrast, the self-appointed destructive critic's role now includes public shaming and blaming, often implying dishonesty on the part of the target and other innuendo based on unchecked assumptions. Targets often seem to be chosen for scientifically irrelevant reasons: their contrary opinions, professional prominence, or career-stage vulnerability.

The destructo-critics are ignoring ethical rules of conduct because they circumvent constructive peer review: They attack the person, not just the work; they attack publicly, without quality controls; they have sent their unsolicited, unvetted attacks to tenure-review committees and public-speaking sponsors; they have implicated targets' family members and advisors. Not all self-appointed critics behave unethically, and some do so more than others. One hopes that these critics aim to improve the field, not harm people. But the fact is that these vigilante critiques are harming people. They are a far cry from peer-reviewed critiques, which serve science without destroying lives.

Ultimately, science is a community, and we are in it together. We agree to abide by scientific standards, ethical norms, and mutual respect. We trust but verify, and science improves in the process. Psychological science has achieved much through collaboration, but also through responding to constructive adversaries who make their critiques respectfully. The key word here is *constructive*.

APS Past President Susan T. Fiske is Eugene Higgins Professor of Psychology and Professor of Public Affairs at Princeton University. Her research focuses on how stereotyping, prejudice, and discrimination are influenced by cooperation, competition, power, and other social relationships. She is a recipient of the 2017 APS James McKeen Cattell Fellow Award.

In short, Fiske doesn't like when people use social media to publish negative comments on published research. She's implicitly following what I've sometimes called the research incumbency rule: that, once an article is published in some approved venue, it should be taken as truth. I've written [elsewhere](#) on my problems with this attitude—in short, (a) many published papers are clearly in error, which can often be seen just by internal examination of the claims and which becomes even clearer following unsuccessful replication, and (b) publication itself is such a crapshoot that it's a statistical error to draw a bright line between published and unpublished work.

Clouds roll in from the north and it started to rain

To understand Fiske's attitude, it helps to realize *how fast* things have changed. As of five years ago—2011—the replication crisis was barely a cloud on the horizon.

Here's what I see as the timeline of important events:

1960s-1970s: Paul Meehl [argues](#) that the standard paradigm of experimental psychology doesn't work, that "a zealous and clever investigator can slowly wend his way through a tenuous nomological network, performing a long series of related experiments which appear to the uncritical reader as a fine example of 'an integrated research program,' [without ever once refuting or corroborating so much as a single strand of the network.](#)"

Psychologists all knew who Paul Meehl was, but they pretty much ignored his warnings. For example, Robert Rosenthal wrote an influential paper on the "file drawer problem" but if anything this distracts from the larger problems of the [find-statistical-significance-any-way-you-can-and-declare-victory paradigm.](#)

1960s: Jacob Cohen studies statistical power, spreading the idea that design and data collection are central to good research in psychology, and culminating in his book, *Statistical Power Analysis for the Behavioral Sciences*. The research community incorporates Cohen's methods and terminology into its practice but sidesteps the most important issue by drastically overestimating real-world effect sizes.

1971: Tversky and Kahneman write "[Belief in the law of small numbers](#)," one of their first studies of persistent biases in human cognition. This early work focuses on researchers' misunderstanding of uncertainty and variation (particularly but not limited to p-values and statistical significance), but they and their colleagues soon move into more general lines of inquiry and don't fully recognize the implication of their work for research practice.

1980s-1990s: Null hypothesis significance testing becomes increasingly [controversial](#) within the world of psychology. Unfortunately this was framed more as a methods question than a research question, and I think the idea was that research protocols are just fine, all that's needed was a tweaking of the analysis. I didn't see general airing of Meehl-like conjectures that much published research was useless.

2006: I first [hear about](#) the work of Satoshi Kanazawa, a sociologist who published a series of papers with provocative claims ("Engineers have more sons, nurses have more daughters," etc.), each of which turns out to be based on some statistical error. I was of course already aware that statistical errors exist, but I hadn't fully come to terms with the idea that this particular research

program, and others like it, were dead on arrival because of too low a signal-to-noise ratio. It still seemed a problem with statistical analysis, to be resolved one error at a time.

2008: Edward Vul, Christine Harris, Piotr Winkielman, and Harold Pashler write a controversial [article](#), “Voodoo correlations in social neuroscience,” arguing not just that some published papers have technical problems but also that these statistical problems are distorting the research field, and that many prominent published claims in the area are not to be trusted. This is moving into Meehl territory.

2008 also saw the start of the blog Neuroskeptic, which started with the usual soft targets (prayer studies, vaccine deniers), then started to criticize science hype (“I’d like to make it clear that I’m not out to criticize the paper itself or the authors . . . I think the data from this study are valuable and interesting – to a specialist. What concerns me is the way in which this study and others like it are reported, and indeed the fact that they are reported as news at all”), but soon moved to larger criticisms of the field. I don’t know that the Neuroskeptic blog per se was such a big deal but it’s symptomatic of a larger shift of science-opinion blogging away from traditional political topics toward internal criticism.

2011: Joseph Simmons, Leif Nelson, and Uri Simonsohn publish a paper, “[False-positive psychology](#),” in *Psychological Science* introducing the useful term “[researcher degrees of freedom](#).” Later they come up with the term [p-hacking](#), and Eric Loken and I speak of [the garden of forking paths](#) to describe the processes by which researcher degrees of freedom are employed to attain statistical significance. ~~The paper by Simmons et al. is also notable in its punning title, not just questioning the claims of the subfield of positive psychology but also mocking it.~~ (Correction: Uri emailed to inform me that their paper actually had nothing to do with the subfield of positive psychology and that they intended no such pun.)

That same year, Simonsohn also publishes a [paper](#) shooting down the dentist-named-Dennis paper, not a major moment in the history of psychology but important to me because that was a paper whose conclusions I’d uncritically accepted when it had come out. I too had been unaware of the fundamental weakness of so much empirical research.

2011: Daryl Bem publishes his article, “Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect,” in a top journal in psychology. Not too many people thought Bem had discovered ESP but there was a general impression that his work was basically solid, and thus this was presented as a concern for psychology research. For example, the New York Times [reported](#):

The editor of the journal, Charles Judd, a psychologist at the University of Colorado, said the paper went through the journal's regular review process. "Four reviewers made comments on the manuscript," he said, "and these are very trusted people."

In retrospect, Bem's paper had **huge, obvious multiple comparisons problems**—the editor and his four reviewers just didn't know what to look for—but back in 2011 we weren't so good at noticing this sort of thing.

At this point, certain earlier work was seen to fit into this larger pattern, that **certain methodological flaws in standard statistical practice were not merely isolated mistakes or even patterns of mistakes, but that they could be doing serious damage to the scientific process**. Some relevant documents here are John Ioannidis's 2005 paper, "Why most published research findings are false," and Nicholas Christakis's and James Fowler's paper from 2007 claiming that obesity is contagious. Ioannidis's paper is now a classic, but when it came out I don't think most of us thought through its larger implications; the paper by Christakis and Fowler is **no longer** being taken seriously but back in the day it was a big deal. My point is, these events from 2005 and 2007 fit into our storyline but were not fully recognized as such at the time. It was Bem, perhaps, who kicked us all into the realization that bad work could be the rule, not the exception.

So, as of early 2011, there's a sense that something's wrong, but it's not so clear to people *how* wrong things are, and observers (myself included) remain unaware of the ubiquity, indeed the obviousness, of **fatal multiple comparisons problems in so much published research**. Or, I should say, the deadly combination of weak theory being supported almost entirely by **statistically significant results which themselves are the product of uncontrolled researcher degrees of freedom**.

2011: Various episodes of **scientific misconduct** hit the news. Diederik Stapel is kicked out of the psychology department at Tilburg University and **Marc Hauser leaves the psychology department at Harvard**. These and other episodes bring attention to the **Retraction Watch** blog. I see a **connection** between scientific fraud, sloppiness, and plain old incompetence: in all cases I see researchers who are true believers in their hypotheses, which in turn are vague enough to support any evidence thrown at them. Recall **Clarke's Law**.

2012: Gregory Francis publishes "**Too good to be true**," leading off a series of papers arguing that **repeated statistically significant results** (that is, standard practice in published psychology papers) **can be a sign of selection bias**. PubPeer starts up.

2013: Katherine Button, John Ioannidis, Claire Mokrysz, Brian Nosek, Jonathan Flint, Emma Robinson, and Marcus Munafo publish the article, “[Power failure: Why small sample size undermines the reliability of neuroscience](#),” which closes the loop from Cohen’s power analysis to Meehl’s more general despair, with the connection being selection and overestimates of effect sizes.

Around this time, people [start](#) sending me bad papers that make extreme claims based on weak data. The first might have been the one on ovulation and voting, but then we get ovulation and clothing, fat arms and political attitudes, and all the rest. The term “Psychological-Science-style research” enters the lexicon.

Also, the replication movement gains steam and a series of high-profile failed replications come out. First there’s the entirely unsurprising lack of replication of Bem’s ESP work—Bem himself wrote a paper claiming successful replication, but his meta-analysis included [various](#) studies that were not replications at all—and then came the unsuccessful replications of embodied cognition, [ego depletion](#), and various other respected findings from social psychology.

2015: Many different concerns with research quality and the scientific publication process converge in the “[power pose](#)” research of Dana Carney, Amy Cuddy, and Andy Yap, which received adoring media coverage but which suffered from the now-familiar problems of [massive uncontrolled researcher degrees of freedom](#) (see this [discussion](#) by Uri Simonsohn), and which failed to reappear in a [replication](#) attempt by Eva Ranehill, Anna Dreber, Magnus Johannesson, Susanne Leiberg, Sunhae Sul, and Roberto Weber.

Meanwhile, the prestigious Proceedings of the National Academy of Sciences (PPNAS) gets into the game, publishing [really bad, fatally flawed papers](#) on media-friendly topics such as [himmicanes](#), [air rage](#), and “[People search for meaning when they approach a new decade in chronological age](#).” These particular articles were all edited by “Susan T. Fiske, Princeton University.” Just when the news was finally getting out about researcher degrees of freedom, statistical significance, and the perils of low-power studies, PPNAS jumps in. Talk about bad timing.

2016: Brian Nosek and others organize a large collaborative replication project. Lots of prominent studies don’t replicate. The replication project gets lots of attention among scientists and in the news, [moving](#) psychology, and maybe [scientific research, down a notch when it comes to public trust](#). There are some rearguard attempts to pooh-pooh the failed replication but they are [not](#) convincing.

Late 2016: We have now reached the “emperor has no clothes” phase. When seemingly solid findings in social psychology turn out not to replicate, we’re no longer surprised.

Rained real hard and it rained for a real long time

OK, that was a pretty detailed timeline. But here’s the point. Almost nothing was happening for a long time, and even after the first revelations and theoretical articles you could still ignore the crisis if you were focused on your research and other responsibilities. Remember, as late as 2011, even Daniel Kahneman was saying of priming studies that “disbelief is not an option. The results are not made up, nor are they statistical flukes. You have no choice but to accept that the major conclusions of these studies are true.”

Then, all of a sudden, the world turned upside down.

If you’d been deeply invested in the old system, it must be pretty upsetting to think about change. Fiske is in the position of someone who owns stock in a failing enterprise, so no wonder she wants to talk it up. The analogy’s not perfect, though, because there’s no one for her to sell her shares to. What Fiske should really do is cut her losses, admit that she and her colleagues were making a lot of mistakes, and move on. She’s got tenure and she’s got the keys to PPNAS, so she could do it. Short term, though, I guess it’s a lot more comfortable for her to rant about replication terrorists and all that.

Six feet of water in the streets of Evangeline

Who is Susan Fiske and why does she think there are methodological terrorists running around? I can’t be sure about the latter point because she declines to say who these terrorists are or point to any specific acts of terror. Her article provides exactly zero evidence but instead gives some uncheckable half-anecdotes.

I first heard of Susan Fiske because her name was attached as editor to the aforementioned PPNAS articles on himmicanes, etc. So, at least in some cases, she’s a poor judge of social science research.

Or, to put it another way, she’s living in 2016 but she’s stuck in 2006-era thinking. Back 10 years ago, maybe I would’ve fallen for the himmicanes and air rage papers too. I’d like to think not, but who knows? Following Simonsohn and others, I’ve become much more skeptical about published research than I used to be. It’s taken a lot of us a lot of time to move to the position where Meehl was standing, fifty years ago.

Fiske's own published work has some issues too. I make no statement about her research in general, as I haven't read most of her papers. What I do know is what Nick Brown [sent me](#):

For an assortment of reasons, I [Brown] found myself reading this article one day: This Old Stereotype: The Pervasiveness and Persistence of the Elderly Stereotype by Amy J. C. Cuddy, Michael I. Norton, and Susan T. Fiske (Journal of Social Issues, 2005). . . .

This **paper was just riddled through with errors**. First off, its main claims were supported by t statistics of 5.03 and 11.14 . . . ummmmm, upon recalculation the values were actually 1.8 and 3.3. So one of the claim wasn't even "statistically significant" (thus, under the rules, was unpublishable).

But that wasn't the worst of it. It turns out that some of the numbers reported in that paper just couldn't have been correct. It's possible that the authors were doing some calculations wrong, for example by incorrectly rounding intermediate quantities. Rounding error doesn't sound like such a big deal, but it can supply a useful set of "degrees of freedom" to allow researchers to get the results they want, out of data that aren't readily cooperating.

There's more at the link. The short story is that Cuddy, Norton, and Fiske made a bunch of data errors—which is too bad, but such things happen—and then **when the errors were pointed out to them, they refused to reconsider anything. Their substantive theory is so open-ended that it can explain just about any result, any interaction in any direction.**

And that's why the authors' **claim** that fixing the errors "does not change the conclusion of the paper" is both ridiculous and all too true. It's ridiculous because one of the key claims is entirely based on a statistically significant p-value that is no longer there. But the claim is true because the real "conclusion of the paper" doesn't depend on any of its details—all that matters is that there's *something*, somewhere, that has p less than .05, because that's enough to make **publishable, promotable claims about "the pervasiveness and persistence of the elderly stereotype" or whatever else they want to publish that day.**

When the authors protest that none of the errors really matter, it makes you realize that, in these projects, the data hardly matter at all.

Why do I go into all this detail? Is it simply mudslinging? Fiske attacks science reformers, so science reformers slam Fiske? No, that's not the point. The issue is not Fiske's data processing errors or her poor judgment as journal editor; rather, what's relevant here is that she's **working within a dead paradigm**. A paradigm that should've been dead back in the 1960s when Meehl was writing on all this, but **which in the wake of Simonsohn, Button et al., Nosek et al., is certainly dead today.** It's the paradigm of the open-ended theory, of publication in top journals

and promotion in the popular and business press, based on “p less than .05” results obtained using abundant researcher degrees of freedom. It’s the paradigm of the theory that in the words of sociologist Jeremy Freese, is “more vampirical than empirical—unable to be killed by mere data.” It’s the paradigm followed by Roy Baumeister and John Bargh, two prominent social psychologists who were on the wrong end of some replication failures and just can’t handle it.

I’m not saying that none of Fiske’s work would replicate or that most of it won’t replicate or even that a third of it won’t replicate. I have no idea; I’ve done no survey. I’m saying that the approach to research demonstrated by Fiske in her response to criticism of that work of hers is an style that, ten years ago, was standard in psychology but is not so much anymore. So again, her discomfort with the modern world is understandable.

Fiske’s collaborators and former students also seem to show similar research styles, favoring flexible hypotheses, proof-by-statistical-significance, and an unserious attitude toward criticism.

And let me emphasize here that, yes, statisticians can play a useful role in this discussion. If Fiske etc. really hate statistics and research methods, that’s fine; they could try to design transparent experiments that work every time. But, no, *they’re* the ones justifying their claims using p-values extracted from noisy data, *they’re* the ones rejecting submissions from PPNAS because they’re not exciting enough, *they’re* the ones who seem to believe just about anything (e.g., the claim that women were changing their vote preferences by 20 percentage points based on the time of the month) if it has a “p less than .05” attached to it. If that’s the game you want to play, then methods criticism is relevant, for sure.

The river rose all day, the river rose all night

Errors feed upon themselves. Researchers who make one error can follow up with more. Once you don’t really care about your numbers, anything can happen. Here’s a particularly horrible example from some researchers whose work was questioned:

Although 8 coding errors were discovered in Study 3 data and this particular study has been retracted from that article, as I show in this article, the arguments being put forth by the critics are untenable. . . . Regarding the apparent errors in Study 3, I find that removing the target word stems SUPP and CE do not influence findings in any way.

Hahaha, pretty funny. Results are so robust to 8 coding errors! Also amusing that they retracted Study 3 but they still can’t let it go. See also [here](#).

I’m reminded of the notorious “gremlins” paper by Richard Tol which ended up having almost as many error corrections as data points—no kidding!—but none of these corrections was

enough for him to change his conclusion. It's almost as if he'd decided on that ahead of time. And, hey, it's fine to do purely theoretical work, but then no need to distract us with data.

Some people got lost in the flood

Look. I'm not saying these are bad people. Sure, maybe they cut corners here or there, or make some mistakes, but those are all technicalities—at least, that's how I'm guessing they're thinking. For Cuddy, Norton, and Fiske to step back and think that maybe almost everything they've been doing for years is all a mistake . . . that's a big jump to take. Indeed, they'll probably never take it. All the incentives fall in the other direction.

In her article that was my excuse to write this long post, Fiske expresses concerns for the careers of her friends, careers that may have been damaged by public airing of their research mistakes. Just remember that, for each of these people, there may well be three other young researchers who were doing careful, serious work but then didn't get picked for a plum job or promotion because it was too hard to compete with other candidates who did sloppy but flashy work that got published in *Psych Science* or *PPNAS*. It goes both ways.

Some people got away alright

The other thing that's sad here is how Fiske seems to have felt the need to compromise her own principles here. She deplores “unfiltered trash talk,” “unmoderated attacks” and “adversarial viciousness” and insists on the importance of “editorial oversight and peer review.” According to Fiske, criticisms should be “most often in private with a chance to improve (peer review), or at least in moderated exchanges (curated comments and rebuttals).” And she writes of “scientific standards, ethical norms, and mutual respect.”

But Fiske expresses these views in an unvetted attack in an unmoderated forum with no peer review or opportunity for comments or rebuttals, meanwhile referring to her unnamed adversaries as “methodological terrorists.” Sounds like unfiltered trash talk to me. But, then again, I haven't seen Fiske on the basketball court so I really have no idea what she sounds like when she's *really* trash talkin'.

I bring this up not in the spirit of gotcha, but rather to emphasize what a difficult position Fiske is in. She's seeing her professional world collapsing—not at a personal level, I assume she'll keep her title as the Eugene Higgins Professor of Psychology and Professor of Public Affairs at Princeton University for as long as she wants—but her work and the work of her friends and colleagues is being questioned in a way that no one could've imagined ten years ago. It's scary,

and it's gotta be a lot easier for her to blame some unnamed "terrorists" than to confront the gaps in her own understanding of research methods.

To put it another way, Fiske and her friends and students followed a certain path which has given them fame, fortune, and acclaim. Question the path, and you question the legitimacy of all that came from it. And that can't be pleasant.

The river have busted through clear down to Plaquemines

Fiske is annoyed with social media, and I can understand that. She's sitting at the top of traditional media. She can publish an article in the APS Observer and get all this discussion without having to go through peer review; she has the power to approve articles for the prestigious Proceedings of the National Academy of Sciences; work by herself and her colleagues is featured in national newspapers, TV, radio, and even Ted talks, or so I've heard. Top-down media are Susan Fiske's friend. Social media, though, she has no control over. That's must be frustrating, and as a successful practitioner of traditional media myself (yes, I too have published in scholarly journals), I too can get annoyed when newcomers circumvent the traditional channels of publication. People such as Fiske and myself spend our professional lives building up a small fortune of coin in the form of publications and citations, and it's painful to see that devalued, or to think that there's another sort of scrip in circulation that can buy things that our old-school money cannot.

But let's forget about careers for a moment and instead talk science.

When it comes to pointing out errors in published work, social media have been *necessary*. There just has been no reasonable alternative. Yes, it's sometimes possible to publish peer-reviewed letters in journals criticizing published work, but it can be a huge amount of effort. Journals and authors often apply massive resistance to bury criticisms.

There's also [this](#) discussion which is kinda relevant:

What do I like about blogs compared to journal articles? First, blog space is unlimited, journal space is limited, especially in high-profile high-publicity journals such as Science, Nature, and PPNAS. Second, in a blog it's ok to express uncertainty, in journals there's the norm of certainty. On my blog, I was able to openly discuss various ideas of age adjustment, whereas in their journal article, Case and Deaton had nothing to say but that their numbers "are not age-adjusted within the 10-y 45-54 age group." That's all! I don't blame Case and Deaton for being so terse; they were following the requirements of the journal, which is to provide minimal explanation and minimal exploration. . . . over and over again, we're seeing journal article, or journal-article-followed-by-press-interviews, as discouraging data exploration and discouraging the expression

of uncertainty. . . . The norms of peer reviewed journals such as PPNAS encourage presenting work with a facade of certainty.

Again, the goal here is to do good science. It's hard to do good science when mistakes don't get flagged and when you're supposed to act as if you've always been right all along, that any data pattern you see is consistent with theory, etc. It's a problem for the authors of the original work, who can waste years of effort chasing leads that have already been discredited, it's a problem for researchers who follow up on erroneous work, and it's a problem for other researchers who want to do careful work but find it difficult to compete in a busy publishing environment with the authors of flashy, sloppy exercises in noise mining that have made "Psychological Science" (the journal, not the scientific field) into a punch line.

It's fine to make mistakes. I've published work myself that I've had to retract, so I'm hardly in a position to slam others for sloppy data analysis and lapses in logic. And when someone points out my mistakes, I thank them. I don't label corrections as "ad hominem smear tactics"; rather, I take advantage of this sort of unsolicited free criticism to make my work better. (See [here](#) for an example of how I adjusted my research in response to a critique which was not fully informed and kinda rude but still offered value.) I recommend Susan Fiske do the same.

Six feet of water in the streets of Evangeline

To me, the saddest part of Fiske's note is near the end, when she writes, "Psychological science has achieved much through collaboration but also through responding to constructive adversaries . . ." Fisk emphasizes "constructive," which is fine. We may have different definitions of what is constructive, but I hope we can all agree that it is constructive to point out mistakes in published work and to perform replication studies.

The thing that saddens me is Fiske's characterization of critics as "adversaries." I'm not an adversary of psychological science! I'm not even an adversary of low-quality psychological science: we often learn from our mistakes and, indeed, in many cases it seems that we can't really learn *without* first making errors of different sorts. What I *am* an adversary of, is people not admitting error and studiously looking away from mistakes that have been pointed out to them.

If Kanazawa did his Kanazawa thing, and the power pose people did their power-pose thing, and so forth and so on, I'd say, Fine, I can see how these things were worth a shot. But when statistical design analysis shows that this research is impossible, or when replication failures show that published conclusions were mistaken, then damn right I expect you to move forward, not keep doing the same thing over and over, and insisting you were right all along. Cos that

ain't science. Or, I should say, it's a really really inefficient way to do science, for individual researchers to devote their careers to dead ends, just cos they refuse to admit error.

We learn from our mistakes, but only if we recognize that they *are* mistakes. Debugging is a collaborative process. If you approve some code and I find a bug in it, I'm not an adversary, I'm a collaborator. If you try to paint me as an "adversary" in order to avoid having to correct the bug, that's *your* problem.

They're tryin' to wash us away, they're tryin' to wash us away

Let me conclude with a key disagreement I have with Fiske. She prefers moderated forums where criticism is done in private. I prefer **open discussion**. Personally I am not a fan of Twitter, where the space limitation seems to encourage snappy, often adversarial exchanges. I like blogs, and blog comments, because we have **enough space** to fully explain ourselves and to give full references to what we are discussing.

Hence I am posting this on our blog, where anyone has an opportunity to respond. That's right, *anyone*. Susan Fiske can respond, and so can anyone else. Including lots of people who have an interest in psychological science but *don't* have the opportunity to write non-peer-reviewed articles for the APS Observer, who *aren't* tenured professors at major universities, etc. **This is open discussion, it's the opposite of terrorism. And I think it's pretty ridiculous that I even have to say such a thing which is so obvious.**

P.S. [More here](#): Why is the scientific replication crisis centered on psychology?