

PAGES

- How you can support Retraction Watch
- Meet the Retraction Watch staff
  - About Adam Marcus
  - About Ivan Oransky
- Papers that cite Retraction Watch
- Privacy policy
- Retracted coronavirus (COVID-19) papers
- Retraction Watch Database User Guide
  - Retraction Watch Database User Guide Appendix A: Fields
  - Retraction Watch Database User Guide Appendix B: Reasons
  - Retraction Watch Database User Guide Appendix C: Article Types
  - Retraction Watch Database User Guide Appendix D: Changes
- The Center For Scientific Integrity
  - Board of Directors
- The Retraction Watch FAQ, including comments policy
  - The Retraction Watch Transparency Index
- The Retraction Watch Leaderboard
  - Top 10 most highly cited retracted papers
- What people are saying about Retraction Watch

AlexS on Medical journal retracts article on "tribalism" after readers call it offensive

Anthony R. Pratkanis on Social psychology in the age of retraction

Adede on PNAS bans author for refusing to share algae strain

ARCHIVES

Select Month

# Social psychology in the age of retraction



Augustine Brannigan

We're pleased to present an excerpt from chapter 10, "The Replication Crisis," of Augustine Brannigan's *The Use and Misuse of the Experimental Method in Social Psychology* (Routledge 2021), with permission from the publisher.

Contemporary social psychology has been seized over the past years by a loss of credibility and self-confidence associated with scientific fraud and unsuccessful attempts to replicate the modern corpus of knowledge. The most notorious case was that of Dietrick Stapel. Fifty-eight papers published over a decade and a half were retracted due to fraud and suspicious research practices.

One of the most poignant questions raised by the review committees in three universities where he worked was how it was possible for such dubious scientific practices to escape the notice of all the academic reviewers in the high-profile journals, the funding agencies and at the scientific conferences. Many statistical anomalies were identified readily by statisticians who assisted in the review of Stapel's papers. The committees were forced to conclude that "there is a general culture of careless, selective and uncritical handling of research and data. The observed flaws were not minor 'normal' imperfections in statistical processing, or experimental design and execution, but violations of fundamental rules of proper scientific research." The culture contributed to the absence of skepticism about Stapel's extraordinary findings.

We tend to think that there is a sharp line between outright fraud and the "massaging" of data. Stapel and Smeesters (another Dutch psychologist implicated in data manipulation) did both, but that part of their publications in which they engaged in grey-area data manipulation appears to be common. The Netherlands committees of inquiry into Stapel were told that "this is what I learned in practice; everyone in my research environment does the same, as does everyone we talk to at international conferences.". Seemsters reported similarly about the generality of data massaging in his area.

John, Lowenstein and Prelec examined questionable research practices in a more general way. They conducted an on-line survey sent to nearly 6000 researchers, including over 2000 psychologists, to estimate the prevalence of the use of self-reported questionable research practices (QRPs). What did they learn from the psychologists? One in ten respondents admitted to having falsified data, 67% reported they selectively reported results that "worked", 74% failed to report all their actual dependent variables, 71% reported that they continued to collect data until they achieved a significant result, 54% reported unexpected findings as having been hypothesized beforehand, and 58% excluded data to enhance the significance of their findings. The highest levels of self-admissions of QRPs were found among social psychologists (40%), followed by cognitive scientists (37%) and neuroscientists (35%).

After his key studies of psychic powers failed to replicate, social psychologist Daryl Bem commented: "If you looked at all my past experiments, they were always rhetorical devices. I gathered data to show how my point would be made. I used data as a point of persuasion, and I never really worried about, 'will this replicate or will this not?'" That view does not appear to be out of line among psychologists with the high level of QRPs identified in the John et al survey. And it reflects the approach in classical social psychology where the experiments of Sherif, Milgram, Asch and Zimbardo were designed as demonstrations, undertaken without explicit hypotheses, control groups or tests of significance. In classical social psychesology (e.g. Milgram, Zimbardo, Sherif), verification bias was legion.

Today there is a heightened level of concern both among scientists as well the educated public with the failures to replicate important experimental work in psychology. The Center for Open Science has initiated important and unprecedented attempts to replicate contemporary research involving scores of colleagues. This has occurred at the same time that one of the most provocative new developments in the field associated with experimental social psychology – priming – has experienced significant levels of failures to replicate. Many of the "instant classics" in this field have been retracted.

Attempts to replicate these studies have sometimes been accompanied by acrimony since the failure to replicate may point on the one side to charges of unprofessional behavior, potential fraud or QRPs, or on the other side to the incompetence, envy and bullying behavior of the replicators. In the age of Retraction Watch, there is no agreed protocol for how replications ought to be undertaken and evaluated. Kahneman proposed an "etiquette" under which the proposed replications would be outlined to the original authors who would have an opportunity to assess the fairness of the replications; the entire correspondence would be transparent and on the record.

Also, in an open email letter circulated widely to colleagues in social psychology after classic studies in priming failed to replicate, Kahneman warned colleagues that that the replication crisis was undermining the credibility of research: "Questions have been raised about the robustness of priming results. . . Your field is now the poster child for doubts about the integrity of psychological research." He went on to suggest that the situation was a looming "train wreck." Others have suggested that that train wreck has already occurred as major studies in social priming research have already failed to replicate.

There are a number of reliability problems in the literature of any field of empirical inquiry that can lead to difficulties in replication. Publication bias is the fact that journals typically do not publish negative findings. Experiments that fail to establish any significant outcomes end up in "the file drawer." Researchers typically do not know what is in their colleagues' file drawers and may undertake research that has already proven fruitless. Publication bias is especially problematic when someone undertakes a meta-analysis that pulls together all the studies of the same problem but only employs what successfully was accepted for publication (due to significance) and overlooks all the negative findings in all the file drawers.

Verification bias is a more serious phenomenon. It refers to a stubborn resistance to accepting the null hypothesis – the assumption that there is no inherent relationship between the variables being studied. The null hypothesis is the default position in experiments. This is what the researcher is attempting to eliminate through experimental investigation. For example, continuing to repeat an experiment until it "works" as desired, or excluding inconvenient cases or results may make the hypothesis immune to the facts. Verification bias amounts to the repression of negative results.

For example, a researcher may exclude some cases because the individuals did not seem to respond to the treatment, or because they were outliers, thus reducing variance in the dependent variable and making statistical significance more likely to emerge. Or a subgroup is selected for analysis because in retrospect, this is the group that yields the significant tests. HARKing – hypothesizing after the results are known – is the re-construction of the objective of the work because some finding reaches statistical significance, even if it was not the objective of the work in the first place. Hence a random event – a false positive – may be treated as a bona fide achievement and motivate others to replicate it.

All these practices have been referred to as "the researchers' degrees of freedom", meaning unjustifiable flexibility in data analysis that is undisclosed to the reader, i.e. employing several different measures of the dependent variable, or controlling for gender effects after the fact to determine if the effect is gender specific. Likewise p-hacking and cherry picking results lead to the same problematic consequences: a large portion of the published literature consists of false positives – studies whose minimum statistical acceptability has been inflated by seemingly minor adjustments to the data which pushes the test just across the .05 statistical threshold.

Remedies to these problems are becoming more widely apparent. They include pre-registration of research proposals at such places as The Center for Open Science and the identification of hypotheses, projected sample size and composition and approach to statistical analysis a priori. Negative findings are becoming somewhat more acceptable in journals, and their publication is encouraged in sources such as the PsychFileDrawer.org. And researchers are being advised to preprint their summary tables on their own websites prior to journal publication. These reforms to research practices promise to make the research process more transparent and credible.

However, this is not the only issue raised in the case of social psychology. In *The Use and Misuse of the Experiment in Social Psychology*, I explore another consideration. Many of the regularities that interest social psychologists actually may not have recurrent law-like properties: How do norms occur? Why are institutions demoralizing? What makes normal people become mass murderers? Hence experiments may not be the optimum method for studying them. In addition, experiments on human subjects are notoriously difficult to standardize, as the recent wave of archival inquiries into Milgram, Sherif, and Zimbardo have demonstrated.

One of the founders of social psychology, Kurt Lewin, described the two great traditions in the history of psychology — the one owed to Galileo, and the other to Aristotle. The former is precise and universalizing while the latter is anthropomorphic and inexact. There are good reasons to believe that the field of social psychology attempted to ground its scientific credibility in Galilean, i.e. experimental, methods. But that approach has led substantively to what John Greenwood has called "the disappearance of the social in social psychology."

Social psychology has become increasingly derivative of individual and cognitive psychology, and the study of emergent social processes between individuals and groups has been replaced with de-contextualized mechanisms. Things such as cognitive dissonance and priming are more easily isolated in the lab than complex social behavior involving social capital and trust. In that case, failures to replicate may not be due to fraud, QRPs or chance, but to the complexity and spontaneity of social behavior, and the failure to adopt research methods that are sensitive to this.

Augustine Brannigan is professor emeritus of sociology at the University of Calgary.

Like Retraction Watch? You can make a one-time tax-deductible contribution or a monthly tax-deductible donation to support our work, follow us on Twitter, like us on Facebook, add us to your RSS reader, or subscribe to our daily digest. If you find a retraction that's not in our database, you can let us know here. For comments or feedback, email us at team@retractionwatch.com.

SHARE THIS:  
[Email](#) [Facebook](#) [Twitter](#)

RELATED  
**"Just significant" results have been around for decades in psychology -- but have gotten worse: study**  
Last year, two psychology researchers set out to figure out whether the statistical results psychologists were reporting in November 12, 2013 In "daniele fanelli"

**New Dutch psychology scandal? Inquiry calls data manipulation, calls for retraction**  
The University of Amsterdam has called for the retraction of a 2011 paper by two psychology researchers after a school April 29, 2014 In "faked data"

**Retraction appears for paper by social psychologist Dirk Smeesters**  
A paper by Dirk Smeesters — the former Erasmus University social psychology professor investigated for serious ir-October 2, 2012 In "dirk smeesters"

June 9, 2021 | Augustine Brannigan | psychology

## One thought on "Social psychology in the age of retraction"

Anthony R. Pratkanis

June 9, 2021 at 2:39 pm

While I certainly agree with Dr. Brannigan's analysis of recent social psychology findings, especially on priming, and have written about it, his depiction of classic social psychology, especially Asch and Milgram, does not match the historical record in regards to verification bias, designed without hypotheses, comparison groups, and replications.

Asch thought that Sheriff's 1935 work on norms and conformity was not a general phenomenon and occurred only because Sheriff used ambiguous stimuli (the auto-kinetic effect). His hypothesis when he began his research was that people would NOT conform if objective (the lines) stimuli were used. Thus, he had a hypothesis (little or no conformity with lines) that was not verified. One can read various editions of his textbook to see him coming to terms with his findings (and his originally thinking about conformity).

Milgram was a post-doc of Asch and his doctoral dissertation replicated Asch in different societies. The goal of his obedience studies was originally similar to his dissertation — to take his obedience experiment to different countries to identify differences in cultures and conformity. His hypothesis was that Germany would show the highest rates, and thus Milgram would establish the basis for the Holocaust. He also thought that Asch's judgment of lines was a trivial task and that conformity/obedience would be much lower when the task involved something of real consequences – shocking a human being. This suited Milgram's original goal and hypotheses, since a lower response rate on the DV gave more room to find a significant effect in more conforming cultures. Of course, his research did NOT verify his original hypotheses.

After obtaining their results, both Asch and Milgram then engaged in a series of studies replicating their findings and looking at factors that increased and decreased the effect (this is where the comparison groups come in). These replications made clear the underlying phenomenon (for example, less obedience when the authority figure was diminished) and also established the findings as reliable. Asch has been replicated over 110 times and Milgram dozens of times. I found this out when in 1997 Dateline NBC asked me to replicate the Asch experiment. I had no idea if a 40 year old study would replicate; I found almost the same results as Asch. A few years later, ABC asked Jerry Burger to replicate the Milgram study, and, he found results similar to those of the original study.

As for another classic of social psychology, cognitive dissonance, being "more easily isolated in the lab than complex social behavior involving social capital and trust," this again does not match the historical record or the contemporary uses of dissonance theory. In developing dissonance theory, Leon Festinger was responding to two real world incidents – the pattern of rumor transmission after an earthquake in India and Mrs. Keech's failed prophecy of the end of the world. In addition, his original theory was also motivated to understand real cases of rejecting evidence about the dangers of smoking and Gunnar Myrdal's dilemma of American racism. Festinger bottled the phenomenon in the lab, and subsequent research found the conditions under which dissonance is most likely to occur. This research has led to practical interventions (for example, Aronson's research on hypocrisy and AIDS prevention) and is useful, as anyone who has been on Facebook would realize, for understanding real world issues today.

These classic studies are not of the same ilk as priming research, and, indeed, provide a standard for how to conduct scientifically valid social psychological research.

REPLY

### Leave a Reply

Your email address will not be published. Required fields are marked \*

COMMENT

NAME \*

EMAIL \*

WEBSITE

BY USING THIS FORM YOU AGREE WITH THE STORAGE AND HANDLING OF YOUR DATA BY THIS WEBSITE PER THE TERMS OF OUR PRIVACY POLICY: [HTTPS://RETRACTIONWATCH.COM/PRIVACY-POLICY/](https://retractionwatch.com/privacy-policy/) \*

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

PREVIOUS  
**Publisher retracts 20 of a researcher's papers — then asks him to peer review**