Everything that happens once can never happen again. But everything that happens twice will surely happen a third time.

Paulo Coelho, The Alchemist.

EINDHOVEN UNIVERSITY OF TECHNOLOGY

AMMODO SCIENCE AWARD 2023 FOR SOCIAL SCIENCES

Twice in the history of social psychology has there been a crisis of confidence. The first started in the 1960s and lasted until the end of the 1970s.

Elms (1975, p. 967) writes “many social psychologists appear to have lost not only their enthusiasm but also their sense of direction and their faith in the discipline's future. Whether they are experiencing an identity crisis, a paradigmatic crisis, or a crisis of confidence, most seem agreed that a crisis is at hand.”

Quiz time!

“We, and social psychology generally, have undergone a crisis, not simply of confidence, but, more profoundly, of paradigm, of our general form of thought. It was as though the life-giving substance in the air we breathed became insufficient; some gasped and suffocated. It was as though we were fish out of water; some flapped around on dry ground.”

1978

2019

1978

2019

Mills, 1978

We have learned that replication is important, and that when we do them, not all results replicate.

"The history of social psychology illustrates the importance of the replication of findings in that many of its initial results have not been confirmed by later investigations."

1953

2013

1953

2013

Festinger & Katz, 1953

One reason for this is due to research practices that inflate alpha levels of studies reported in the literature.

“By seemingly conservative estimates, the probability that a finding (reported at $p = .05$) is a result of chance is not .05, but is closer to .50. That is, were these estimates approximately correct, they would indicate that about one-half of the original findings reported at this level in behavioral science journals could have resulted solely from chance variations.”

1967

2005

1967

2005

Neher, 1967

But we also have problems with the strength of our theories.

"Social psychology is clearly in need of new and better theories. Probably the most persistent complaint in the field's history, from within and without, is that it is largely empirical, with little theoretical guidance."

1975

2015

1975

2015

Elms, 1975

The work we do can further be criticized for its lack of relevance, or a ‘crisis of relevance’.

“As a science directly concerned with behavioral and social processes, psychology might be expected to provide intellectual leadership in the search for new and better personal and social arrangements. In fact, however, we psychologists have contributed relatively little of real importance —even less than our rather modest understanding of behavior might justify.”

1969

2021

1969

2021

Miller, 1969

We focus too much on a small set of conditions, leading to a generalizability crisis.

"in the event that a result is replicable, there is little likelihood that it will be sufficiently general across minor variations in stimulus conditions to identify scientifically useful relationships."

1980

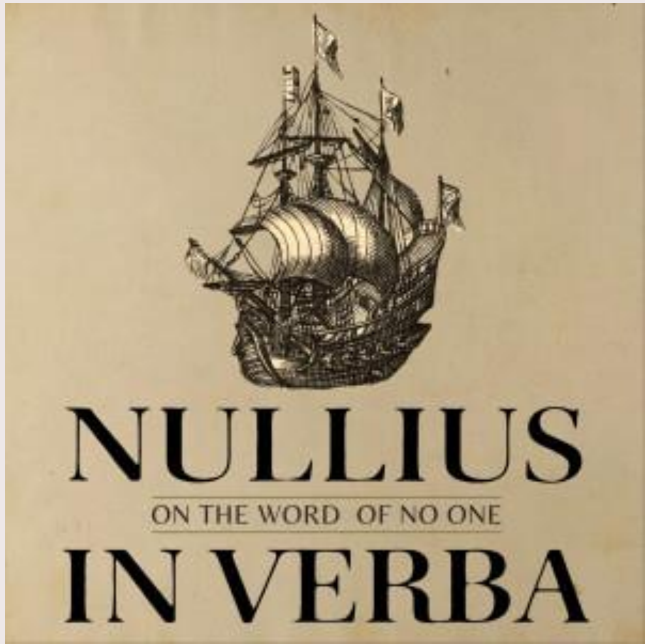
2021

1980

2021

Epstein, 1980

(If you want more of this, listen to our podcast ‘Nullius in Verba’ episode “Episode 16: Vetus Crisi Replicatio”)



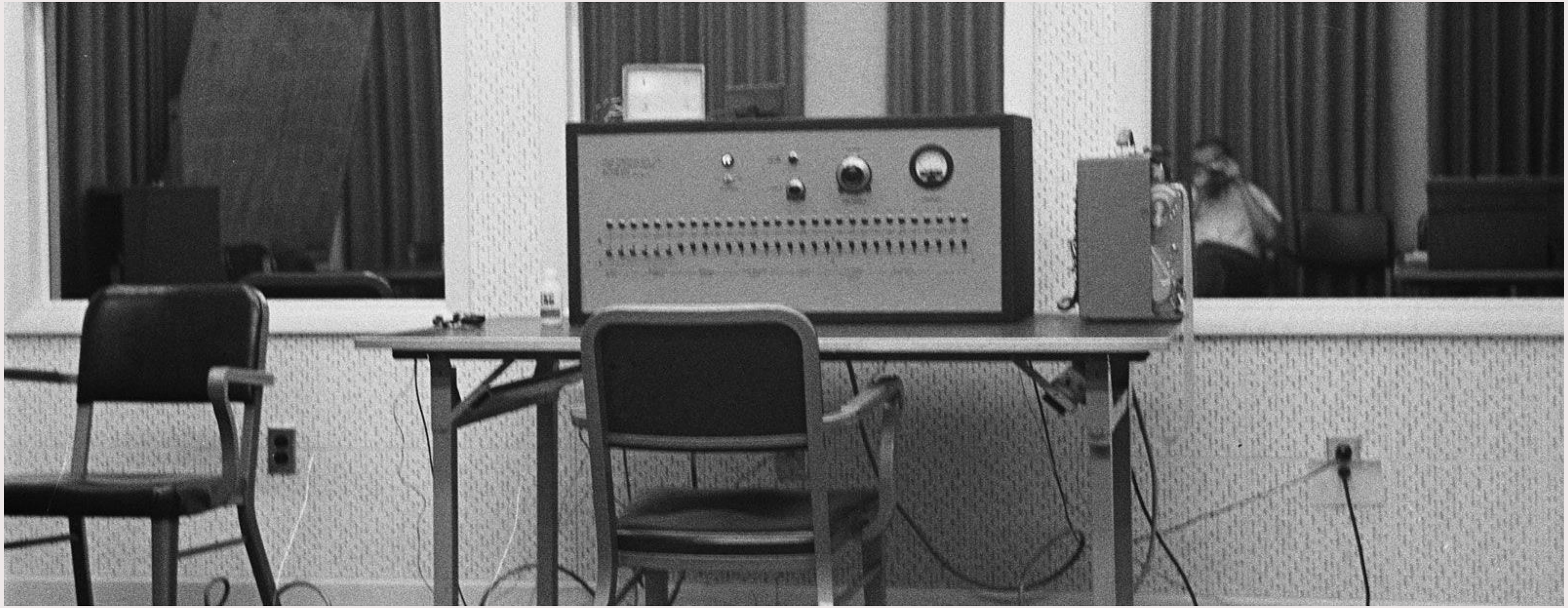
Jones (1985, p. 100) remarks: "The crisis of social psychology has begun to take its place as a minor perturbation in the long history of the social science. The intellectual moment of the field has not been radically affected by crisis proclamations".

And of course, some disagreed there was a crisis at all. Deutsch (1976, p. 134): “Were I to engage in a polemic about theorizing in social psychology, my inclination would be to attack the “doom-criers”, those who assert that social psychology is in a "crisis which it must overcome if it is to survive."“

So did we solve *any* of the problems from the previous crisis?

Yes, we solved concerns about research ethics. It is worth reflecting on how we did this.

Questions were raised regarding the acceptability of deception, and whether the negative consequences for subjects were balanced by the positive consequences of the knowledge that was gained.



The Cook Committee surveyed 18000 APA members, and in 1973 the APA published ‘Ethical Principles in the Conduct of Research with Human Participants’. This led to ethical review boards.

Stark, 2010

The APA took action, because they feared that if they would not, action would be forced upon them.

A replication crisis

“The current incentive structure within psychology gives rise to a social dilemma in which researchers’ collective interests run in the opposite direction as the interests of individual researchers. The research community collectively benefits if researchers take the trouble to replicate each other’s work, because this improves the quality and reputation of the field by showing which observations are reliable. Yet, individually, researchers are better off by conducting only original research, because this will typically yield more publications and citations and thus ultimately greater rewards in terms of better jobs and more grant money.”

Koole & Lakens, 2012

When a finding fails to replicate, logically there are three possible reasons for a non-significant result. First, the original result could be a false positive, or Type 1 error. Second, the replication study yielded a false negative, or a Type 2 error. And third, the replication study might yield a non-significant result because there is an unknown moderator.

Sidman (1960, p. 63), writes: "Failure to replicate a finding, within or among species, is a result of incomplete understanding of the controlling variables. This positive approach, as contrasted to the negative attitude that failure to replicate must brand a process as nongeneral, is actually the only road to an adequate evaluation of generality."

In the second crisis, it became more accepted to attribute failures to replicate to a Type 1 error instead.

There were at least 5 reasons for this.

1) publication of studies that failed to replicate highly cited findings in the field,
2) the attention that these replication failures received
3) a more widespread understanding of research practices that substantially inflate error rates
4) making it more acceptable to interpret original findings as Type 1 errors
5) meta-scientific evidence that the rate at which studies could be replicated was surprisingly low.

A theory crisis

Theoretical retardation. Philosophers of science concur that scientific advances are advances in _theory_. But the state of theory in personality and social psychology may well be reason for concern. A striking illustration of the problem is contained in a recent article by Harris (1974) where popular social psychological theories (notably, Festinger's theory of social comparison processes, the theory of cognitive dissonance and the equity theory) are shown to be riddled by fundamental ambiguities (if not downright inconsistencies) in their formulation. That such conceptualizations (laudable in and of themselves) may have persisted in a badly flawed form for entire decades of widespread mention attests strongly to a lack of serious concern about conceptual issues in a field permeated by unbridled empiricism. Indeed, numerous research papers in social and personality psychology convey the impression that the theoretical issues invoked are little more than an excuse, a thin disguise for the researcher's real objective: the execution of yet another experiment!

Kruglanski, 1975

I do not think that there is any dispute about this matter among psychologists familiar with the history of the other sciences. It is simply a sad fact that in soft psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that *cumulative* character that is so impressive in disciplines like astronomy, molecular biology, and genetics.

Meehl, 1978

What was to blame? 1) mindless statistics.

Another impression given by these multiple exercises in partial theoretical integration is that in the building of our science, we over-value *p*-levels and undervalue the judgmental appraisal of evidence. The

bricks may be culled by *p*-level, but the mortar, the girders, the theoretical structure are necessarily judgmental through and through, with little help from statistics. The linkages are often much looser than we like to pretend. And even the substantiality of our *p*-sorted bricks is suspect, when we know that given enough cases, the null hypothesis rarely prevails in nature. We also know about experimenters' practices in the use of pretests and pilot runs to decide whether a manipulation is appropriate—and about editors' lack of interest in publishing negative results.

Brewer Smith, 1972

What was to blame? 2) 'fun' studies.

Clever experimentation on exotic topics with a zany manipulation seems to be the guaranteed formula for success which, in turn, appears to be defined as being able to effect a *tour de force*.[1] One sometimes gets the impression that an ever-growing coterie of social psychologists is playing (largely for one another's benefit) a game of "can you top this?" Whoever can conduct the most contrived, flamboyant, and mirth-producing experiments receives the highest score on the kudometer. There is, in short, a distinctly exhibitionistic flavor to much current experimentation, while the experimenters themselves often seem to equate notoriety with achievement.

Ring, 1967

Have anything changed? No.

Thus, psychologists (a) lack a collective, coordinated research program on theory formation; (b) are rarely trained to develop skills conducive to theory development; and (c) live in a research culture that endorses the norm that science is defined by its methods of hypothesis testing rather than theory construction more broadly. The central idea of this article is that to break this theoretical stalemate we need a methodology that organizes the practice of theory formation so that it can be developed, practiced, and taught in psychology. To

Borsboom et al., 2021

“Psychologists who are ignorant of intellectual history are condemned to repeat it in their laboratories.

Richard E. Nisbett, The anticreativity letters

A crisis of applicability

As a science directly concerned with behavioral and social processes, psychology might be expected to provide intellectual leadership in the search for new and better personal and social arrangements. In fact, however, we psychologists have contributed relatively little of real importance—even less than our rather modest understanding of behavior might justify. We should have contributed more; although our scientific base for valid contributions is far from comprehensive, certainly more is known than has been used intelligently.

Miller.,1969, Presidential Address APA

Applied Social Psychology: The Unfulfilled Promise[1]

Robert Helmreich
The University of Texas at Austin

Prospects. Predicting the future course and orientation of the field is chancy, especially since the accuracy of psychological forecasting is not markedly superior to that of astrology or haruspicy. However, several outcomes do seem possible. One is that the pursuit of social psychological truths will degenerate into a form of laboratory-based, mental masturbation, valid in its own right, but devoid of contact with mundane reality.

Helmreich,1975

The hope was there would be an increase in field experiments instead of field studies.

But metascientific research shows there is no increase in field studies in response to the first (Higbee & Wells, 1972) or second (Doliński, 2018) crisis.

Quiz time!

What are the benefits of doing field studies?

What are the downsides of doing field studies?

Have concerns about the relevance of psychological science disappeared?

Or have we just stopped talking about it?

A generalizability crisis

If the multitude of social-psychological findings cannot aid the planners of society, it is apparently not because we have been researching the wrong topics. It must be that our data are not generalizable to the objects of our studies in their natural, ongoing states. This is a basic inadequacy of methodology rather than direction, and it will not be resolved by pontifical edicts from any source about what to study and where.

Silverman,1971

that has been conducted to date. The difficulty with the typical laboratory experiment is not that it cannot yield meaningful generalizations, but that there is no way of establishing within the confines of the experiment that it is likely to have done so. At the same time, there is an unfortunate dearth of replication studies. Moreover, as is demonstrated shortly, the very nature of the paradigm of the single-session experiment is such that very few findings, no matter what their level of statistical significance, are apt to be replicable. Further, in the event that a result is replicable, there is little likelihood that it will be sufficiently general across minor variations in stimulus conditions to identify scientifically useful relationships.

Epstein,1980

“At the same time, the current focus on reproducibility and replicability risks distracting us from more important, and logically antecedent, concerns about generalizability. The root problem is that when the manifestation of a phenomenon is highly variable across potential measurement contexts, it simply does not matter very much whether any single realization is replicable or not”.

Yarkoni,2021

A meta-scientific study by Dipboye and Flanagan (1979) found that both field studies and lab experiments sample from a limited range of subject, settings, and behaviors, and that both types of studies therefore provide little guarantees that observed findings will generalize.

The Psychological Science Accelerator aims to accumulate generalizable knowledge.

A methodological crisis

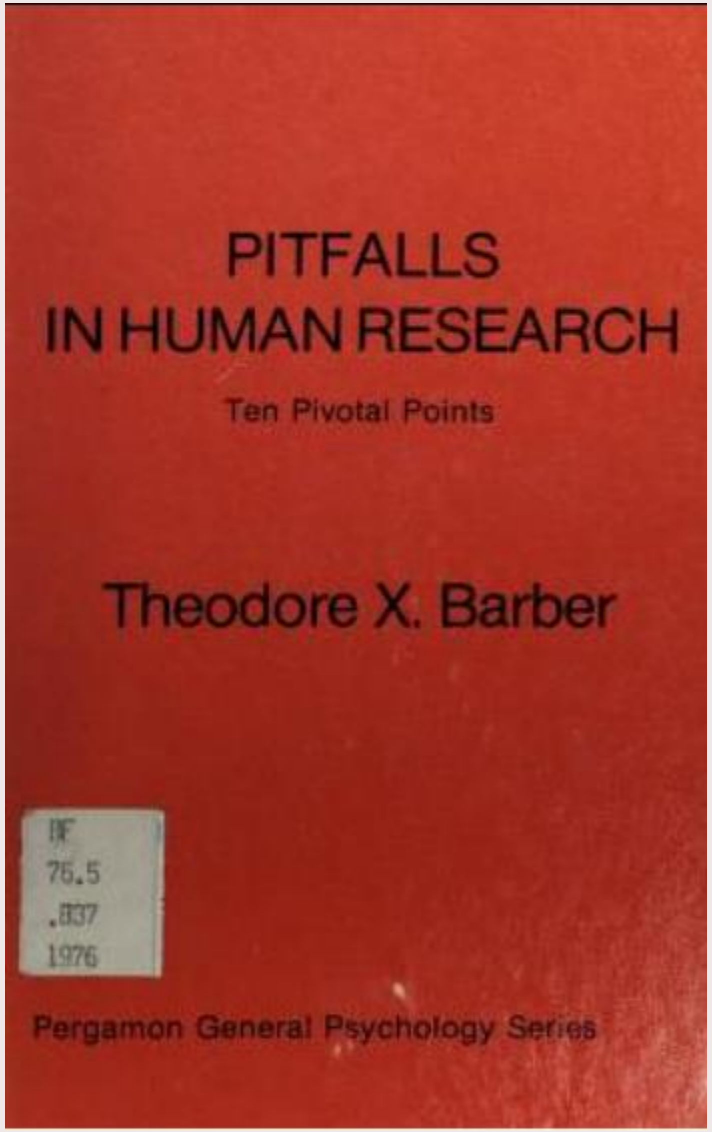
PROBABILITY PYRAMIDING, RESEARCH ERROR AND THE NEED FOR INDEPENDENT REPLICATION[1]

ANDREW NEHER[2]

Research Department, Youth Opportunities Center
San Francisco

Screening of results of behavioral science investigation occurs at several successive stages from the individual analysis to final publication, and selection is partly determined by a low probability that a result could have occurred by chance (i.e., a low *p* level). It is demonstrated that a relatively small degree of such selection is sufficient to pyramid the *p* level to many times its reported .01 or .05 size, resulting in a large proportion of fallacious "findings" in the behavioral science literature. This probability pyramid is one of a class of serious research errors which can be adequately reduced only through the practice of independent replication, the adoption of which is crucial to the behavioral sciences.

1) performing unplanned analyses,
2) hypothesizing after results are known
3) performing a large number of tests and only reporting significant results
4) cutting and slicing data in originally unintended ways
5) not correcting for multiple comparisons
6) selectively reporting significant results,
7) not reporting non-significant results,
8) checking for errors after negative results, but not checking for errors after positive results
9) reporting statistically significant but practically insignificant effects



This is arguably the area where most progress has been made this crisis – high awareness, preregistration, Registered Reports, bias detection.

What causes crises?

Bad incentives.

years. So far as quantity itself goes, the present period, beginning with 1921, is perhaps the most productive epoch in American sociology. But comparison of the product with that of the first decade of this century will reveal a much smaller percentage of notable works. If it be true that, for the time being at least, the quality of American sociological writing is in inverse ratio to its quantity, the reason is to be sought, among other things, in the fact, first, that the system of promotion used in our universities amounts to the warning, "Publish or perish!" In the sec-

Clark,1927

Publication bias.

THE SECRETS WE KEEP

We might better label this game "Dear God, Please Don't Tell Anyone." As the name implies, it incorporates all the things we do to accomplish the aim of looking better in public than we really are.

The most common variant is, of course, the tendency to bury negative results. I only recently became aware of the massive size of this great graveyard for dead studies when a colleague expressed gratification that only a third of his studies "turned out"—as he put it.

Dunnette,1966

A lack of coordination. If we want to solve our remaining crises we will need a coordinated effort – this means shared goals, interdependencies between researchers, and management.

Rasti, S., Vaesen, K., & Lakens, D. (2025). *The Need for Scientific Coordination*. OSF. https://doi.org/10.31234/osf.io/vjcfk_v2

Sajedeh Rasti

Grazi!

https://osf.io/ejqa2/