

RUNNING HEAD: TWO CRISES

**Concerns about Replicability, Theorizing, Applicability, Generalizability, and  
Methodology across Two Crises in Social Psychology**

Daniël Lakens

Eindhoven University of Technology

Correspondence can be addressed to Daniël Lakens, Human Technology Interaction Group,  
ATLAS 9.402, PO Box 513, 5600MB Eindhoven, The Netherlands. E-mail:  
D.Lakens@tue.nl.

## **Abstract**

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, and the second crisis dominated the 2010s. In both these crises, researchers discussed fundamental concerns about the replicability of findings, the strength of theories in the field, the societal relevance of research, the generalizability of effects, and problematic methodological and statistical practices. On the basis of extensive quotes drawn from articles published during both crises, I explore the similarities and differences in discussions across both crises in social psychology.

Twice in the history of social psychology has there been a crisis of confidence. The first started in the 1960's and lasted until the end of the 1970's. Bakan (1967, p. xii) states "I believe there is a crisis in research in psychology", and 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.". The second crisis dominated the 2010's, and almost forty years after the first crisis, Pashler and Wagenmakers (2012, p. 528) wrote "Is there currently a crisis of confidence in psychological science reflecting an unprecedented level of doubt among practitioners about the reliability of research findings in the field? It would certainly appear that there is." What were the similarities and the differences in both crises in social psychology? Which fundamental concerns were raised in each crisis? What were the similarities in the concerns that were raised, and how did concerns differ? Which improvements were introduced in response to each crisis, and which challenges have remained unaddressed, to possibly resurface in a future crisis?

Many researchers have used the term 'crisis' over the last century to discuss what they considered to be fundamental challenges to the field (Sturm & Mülberger, 2012). Deutsch (1976, p. 134) already remarked "The crisis in social psychology is not new; we are in a perpetual crisis." And yet, only twice in psychology's history was the concern among scientists sufficiently widespread, with a continuous discussion in journals in the field by leaders in the field, that lasted for more than a decade. The first crisis started in the 1960's with methodological criticism on the reliability of data collection from human participants due to demand and experimenter effects. The work by Rosenthal (whose 1966 book

‘Experimenter effects in behavioral research’ is still worth reading in full)<sup>1</sup> was an important instigator, which was further amplified by Gergen’s (1973) article ‘Social psychology as history’ which challenged the idea social psychology could produce generalizable knowledge.

The second crisis started in social psychology in 2011 with the publication of an article that provided empirical support for extra-sensory perception (Bem, 2011), which coincided by the publication of a failed replication of a high profile finding in the field (Doyen et al., 2012). These publications raised concerns about research practices that inflated the rate of false positives in the literature. As these concerns about replicability and methodological practices proved relevant for other disciplines as well, the second crisis rapidly spread beyond social psychology. Although scientists in all disciplines regularly raise concerns about how their fields function, I focus on discussions among social psychologists during the first crisis and the second crisis. The fact that history has repeated itself has not gone unnoticed among psychologists writing about the second crisis (Earp & Trafimow, 2015; Giner-Sorolla, 2012; Hales, 2016; Pettigrew, 2018), just as researchers writing during the first crisis recognized historical precedents (Lewin, 1977; Minton, 1984).

During both crises there were diverging views on whether social psychology was in an actual state of crisis. Reflecting on the first crisis, Mills (1979) writes: “We, and social psychology generally, have undergone a crisis, not simply of confidence, as Elms suggests (1975), 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

---

<sup>1</sup> I agree with the analysis of Barber and Silver (1968) that the evidence for experimenter bias presented by Rosenthal is often weak, and perhaps ironically, shows clear signs of inflated Type 1 error rates due to flexibility in how the data is analyzed. Despite this weak evidentiary basis, the impact of these studies was great.

suffocated. It was as though we were fish out of water; some flapped around on dry ground. Although painful, the crisis is serendipitous to the extent that the early translucent form of thought becomes, through its inadequacy, more apparent, both in its productive and counterproductive features and, consequently, more accessible for appreciation and redesign.” Surveying researchers in psychology at the end of the first crisis, Nederhof and Zwier (1983) found widespread agreement that social psychologists had shown great concern about the state of their field in the last decade, and a third of the scientists still believed their field was in a state of intellectual disarray at the end of the first crisis. During the second crisis, a similar survey across scientific disciplines revealed many researchers experienced a crisis of confidence (Baker, 2016).

Beyond these surveys, it is notable that highly successful social psychologists admitted to personally believing the field was in a crisis. In 1970, Berkowitz wrote in personal correspondence to M. B. Smith (1972): “At any rate, it seems to me... that social psychology is now in a “crisis stage,” in the sense that Kuhn used this term in his book *The Structure of Scientific Revolutions*. We seem to be somewhat at a loss for important problems to investigate and models to employ in our research and theory. It is certainly time to take stock, to see where we are and where we should go...”<sup>2</sup> Sherif (1977) writes: “Social psychology in this country is going through an ironic and unsettling state in its development. Ironic and unsettling, because on the surface it is thriving in a sky-rocketing boom of output in research and publication and, at the same time, the ratio of chaff that is piling up is enormous relative

---

<sup>2</sup> It is doubtful any crisis in psychology meets the (arguably vague) requirements of a Kuhnian crisis (Sturm & Mülberger, 2012). Arguably, the question whether a crisis is a Kuhnian crisis is purely academic and has no practical consequences.

to the scanty yield in substance that will survive. This contradictory state of things is at the bottom of both the crisis and the unsettling malaise it arouses.”

Other psychologists agreed there were problems to address, but believed the crisis narrative was overblown. “In this brief response, I want to suggest that the problem is more fundamental (and the solution less practically difficult) than has previously been considered in the burgeoning literature on the "crisis" of social psychology. While I agree with the critics that we are not sufficiently applied, theoretical, field oriented, or historical in our approach, and that some major rethinking is necessary, it also seems that we have made significant progress that can be the basis for real, satisfying success in the future” (Aron, 1979). Deutsch (1976, p. 134) writes “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. "" Others talk about “the alleged ‘crisis’ in social psychology” (Nederhof & Zwier, 1983) or the “so-called "crisis literature."" (Senn, 1988). This use of quotation marks is also common among more critical voices during the second crisis, such as “the so-called “replication crisis” of the 2010s” (Flis, 2019), with Stroebe and Strack (2014, p. 59) writing “the alleged “crisis of replicability” is primarily due to an epistemological misunderstanding that emphasizes the phenomenon instead of its underlying mechanisms.” In this article, I will follow Goertzen’s (2008, p. 831) analysis of the term ‘crisis’, who writes: “I argue they mean there is a fundamental problem that is of serious importance. The adjectives from the previous sentence are significant here. The problem is fundamental — not tangential — to the discipline and its knowledge, and it is of serious — not passing — importance. I believe it is this combination of the problem being fundamental and serious that has led authors to refer to it as constituting a ‘crisis’ — despite the fact that a literal interpretation of that term is imprecise.”

**What is it a crisis of?**

Even when there was widespread agreement that the field was faced with fundamental problems, there is a noticeable pluralism during both the first and the second crisis about what researchers believe to be the causes of the crisis (Sherif, 1977). A crisis seems to provide authors with a useful vehicle to raise issues that they have been dissatisfied with for a long time. Similar themes emerge in both crises, albeit it in different prominence. Researchers debate the replicability of findings, the state of theories in their field, the generalizability of research findings, the practical relevance for society of the research the discipline produces, and a concern about methodological practices. During both crises researchers agree that many of these issues are caused by the incentive structure in academia. Each of these themes will be reviewed in turn, with a focus on similarities and differences between the two crises.

### **Is there a replication crisis?**

In an editorial in the *Journal of Personality and Social Psychology*, Greenwald (1976) writes: “There may be a crisis in personality and social psychology, associated with the difficulty often experienced by researchers in attempting to replicate published work. A precise statement of the magnitude of this problem cannot be made, since most failures to replicate do not receive public report”. A similar concern about the replicability of findings is expressed by Epstein (1980, p. 790): “Not only are experimental findings often difficult to replicate when there are the slightest alterations in conditions, but even attempts at exact replication frequently fail.” Neher (1967, p. 262) concludes: “The general adoption of independent replication as a requirement for acceptance of findings in the behavioral sciences will require the efforts of investigators, readers, and publishing editors alike. It seems clear that such a policy is both long overdue and crucial to the development of a sound body of knowledge concerning human behavior.” Lubin (1957) suggests that, where relevant, manuscripts that demonstrate the replicability of findings should receive a higher publication priority. At the same time, the topic of how replicable published research is was not central to

the first crisis. Few replication studies are performed, and the perceived problem with replicability is mainly technical, in that not sufficient information is provided in published articles to allow peers to perform a close replication study. For example, Pereboom (1971, p. 442) writes: “Related to the above is the common difficulty of communicating all important details of a psychological experiment to one's audience. [...] Investigators attempting to replicate the work of others are painfully aware of these informational gaps.”

A noticeable change in the second crisis is that failures to replicate have more systematically started to appear in the scientific literature, of which the Reproducibility Project: Psychology (Open Science Collaboration, 2015) has received the most attention, but single studies (Doyen et al., 2012; Ritchie et al., 2012) and multi-lab replication studies (Klein et al., 2014; Simons et al., 2014) have been equally important. What has not changed is that the extent of the problem remains difficult to quantify. As was already acknowledged in the Reproducibility Project: Psychology (2015, pp. aac4716-7): “After this intensive effort to reproduce a sample of published psychological findings, how many of the effects have we established are true? Zero. And how many of the effects have we established are false? Zero. Is this a limitation of the project design? No. It is the reality of doing science, even if it is not appreciated in daily practice. Humans desire certainty, and science infrequently provides it. As much as we might wish it to be otherwise, a single study almost never provides definitive resolution for or against an effect and its explanation.” Even though the extent of the replicability problem remains difficult to quantify, awareness of the fact that it is not negligible, at least in certain research areas, is much greater after the second crisis.

During the first crisis there were no attention-grabbing failures to replicate important findings in the field. Researchers mainly pointed out the importance of publishing replication studies. Lykken (1968, p. 159) writes: “Ideally, all experiments would be replicated before publication but this goal is impractical.”, Loevinger (1968, p. 455) makes a similar point:



“Most studies should be replicated prior to publication. This recommendation is particularly pertinent in cases where the results are in the predicted direction, but not significant, or barely so, or only by one-tailed tests.” N. C. Smith (1970, p. 974) notes how replication studies are neglected: “The review of the literature on replication and cross-validation research has revealed that psychologists in both research "disciplines" have tended to ignore replication research. Thus, one cannot help but wonder what the impact might be if every investigator repeated the study which he believed to be his most significant contribution to the field.” Samuelson (1980, p. 623) notes in the specific context of Watson’s ‘Little Albert’ study: “Beyond this apparent failure of internal criticism of the data is another one that is even less debatable: the clear neglect of a cardinal rule of scientific method, that is, replication.”. And yet, no one addresses these concerns by performing direct replications of important findings in the field.

During both crises researchers point out that the incentive structure does not reward replication studies. “Some journals explicitly state that they do not accept replication studies in principle, while others implicitly follow a similar policy; it is not surprising that few are ever published” (Fishman & Neigher, 1982, p. 539). Little changed, much to the dismay of many scientists, as illustrated by the public outcry when at the start of the second crisis failed replications of Bem’s pre-cognition study were desk-rejected by the editor of JPSP, Eliot Smith, who stated “This journal does not publish replication studies, whether successful or unsuccessful” and “We don’t want to be the Journal of Bem Replication” (Aldhous, 2011). Some researchers proposed novel publication formats to solve the problem. For example, Ahlgren (1969, p. 471) discusses a “hypothetical Journal of Replicated Studies in Psychology” which publishes “only those studies that had been successfully replicated (whether the null hypothesis is rejected or not)”, while “In the back of the Journal would be listed abstracts of studies that have been provisionally accepted for publication, contingent on

successful replication” and “Also in the back might be an edifying (and sobering) table of titles of previously listed abstracts, marked "pending replication," "supported by replication," or "confounded by replication.”” Relatedly, Rosenthal (1966, p. 323) states: “In order to benefit properly from replications actually carried out, it is essential that these be routinely published, even if only as brief notes with fuller reports available from the experimenter, from a university library, or from the American Documentation Institute.” Similar calls to publish brief summaries of replication studies have been made in the second crisis: “Replications could be published in a special section of each journal, much like the sections that are currently used for making editorial statements. If journals still appear in print, the print version of a journal issue could contain short abstracts of new replications. The complete manuscripts of replications could be published as supplementary online materials, which become instantly available upon accessing the original report” (Koole & Lakens, 2012, p. 611). None of these recommendations have been implemented. However, an important change during the second crisis is that standalone replication studies have become easier to publish, either as the novel Registered Report publication format (Nosek & Lakens, 2014), or in journals that explicitly state they will consider replication studies for publication. The increase in online open access journals that do not consider novelty as a publication criterion also played a role.

Rosenthal also suggests more systematic approaches to increase the number of replications. “Who, then, on any large scale will provide us with the necessary replications? McGuigan's (1963) data and Woods' (1961) suggest that there are now enough experiments carried out and reported by multiple authors for there to be no hardships in subdividing these studies into as many complete replicates as there are investigators. The total investment of time would not be increased, but the generality of the results would be” (Rosenthal, 1966, p. 323). A similar idea to distribute replications across researchers is proposed during the second

crisis: “However, a potential barrier to independent, prepublication replication attempts is that many researchers have a difficult time finding other labs to conduct such attempts.

StudySwap can be used to find an independent research team to conduct a replication attempt of a not-yet-published study” (Chartier et al., 2018, p. 575).

Another solution to increase the number of direct replication studies that are performed that has been raised repeatedly is to let students perform replication studies. “Why don't psychologists repeat fundamental experiments, as do physicists? This question has popped up in a number of informal discussions in which I happened to participate. [...] Such repeat experiments are very satisfactory for laboratory courses, honors candidates, and even MA theses. They furnish excellent training” (Schlosberg, 1951, p. 177). The same point is made by Frank and Saxe (2012, p. 600) during the second crisis: “Replication is held as the gold standard for ensuring the reliability of published scientific literature. But conducting direct replications is expensive, time-consuming, and unrewarded under current publication practices. So who will do them? The authors argue that students in laboratory classes should replicate recent findings as part of their training in experimental methods.”

Not all replications are equally useful. First of all, there is little use in replicating a badly designed study, as was already noted by Rosenthal (1966, p. 327) during the first crisis: “A very badly done experiment profits us little, upon even many replications.”, and noted for example by Pittelkow and colleagues (2021, p. 212) during the second crisis. “Replication studies for which the original methodology may not have been optimal [...] may not have as much merit as replication studies for which the methodology [...] was comparatively high. Such considerations are important in times where limited funding means replicating every single study may not be feasible.” More contentious is whether we can actually perform a similar enough replication study. During both crises researchers acknowledge it is impossible to perform the same study twice. “But to avoid the not very helpful conclusion that there can

be no replication in the behavioral sciences, we can speak of relative replications. We can order experiments on how close they are to each other in terms of subjects, experimenters, tasks, and situations” (Rosenthal, 1966, p. 321). This issue became much more pertinent when researchers actually started to perform direct replications during the second crisis, and remains an ongoing topic of discussion. A more worked out proposal comes from Lebel and colleagues (2017, p. 255) during the second crisis who note: “Replications lie on an ordered continuum of methodological similarity to an original study, ranging from highly similar to highly dissimilar.” They suggest specifying different aspect of the design (e.g., the operationalization of the independent and dependent variable, the stimuli, the procedure, the context, etc.) and listing the similarity to the original study for each aspect. During both crises, researchers have pointed out there is a role for both conceptual (“constructive”) replication studies and direct (“operational”) replication studies (Crandall & Sherman, 2016; Lykken, 1968; Zwaan et al., 2018), even though there has also been criticism on the idea that we should perform more direct replication studies during both crises (Hunt, 1975; Stroebe & Strack, 2014).

It is far from given that social psychologists can repeat the same study over time. Indeed, the argument put forward by Gergen (1973) that most (if not all) effects in psychology are not stable across time and context was one of the starting points for the first crisis in social psychology. Gergen writes: “It is the purpose of this paper to argue that social psychology is primarily an historical inquiry. Unlike the natural sciences, it deals with facts that are largely nonrepeatable and which fluctuate markedly over time” (1973, p. 310).<sup>3</sup> Gergen does not necessarily dismiss the possibility that some effects turn out to be quite

---

<sup>3</sup> Whether there is actually more variability in the social science than in the physical sciences has been contested (Hedges, 1987).

lasting, but whether effects are stable over time is anything but certain: “We must think, then, in terms of a continuum of historical durability. with phenomena highly susceptible to historical influence at one extreme and the more stable processes at the other.” This work directly inspired the social constructionist movement in psychology (Gergen, 1985). The fact that more replication studies are published during the second crisis made it possible to empirically explore the challenging question of whether replication attempts failed because the original study was a false positive, or because the replication study was not well-designed (Ebersole et al., 2020; Luttrell et al., 2017).

Despite the impact Gergen’s views had on the first crisis, there were also critics of his arguments (e.g., Manis, 1975), and the social constructionist perspective has largely (although not completely) disappeared in the second crisis. The question whether failures to replicate are due to contextual factors continues to be a topic of discussion (Pettigrew, 2018; Van Bavel et al., 2016). An intriguing change during the second crisis is a more pragmatic approach to the question of whether failures to replicate should be attributed to differences in the context. Instead of putting the burden of proof that there are no contextual moderators on the researcher who performs a replication study, the original authors are held accountable for any failure to specify the context to which their finding is expected to generalize. By requiring researchers to specify a ‘Constraints on Generality statement’ (Simons et al., 2014) it now becomes the responsibility of the original researcher to state which contextual moderators they believe would make effects they have observed disappear. Replication studies can now sample intentionally from within or beyond the specified constraints on generality to perform informative tests of the replicability and generalizability of findings.

### **Is there a theory crisis?**

The first crisis was strongly inspired by a general dissatisfaction with the quality of psychological theories, and a perceived lack of progress. For example, Elms (1975, p. 972)

writes “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.” Kruglanski (1975, p. 491) expresses the point in even stronger terms, under the heading “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.” The lack of theoretical progress was famously criticized by Meehl (1978, p. 807): “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.” Similar sentiments were expressed earlier by Pereboom (1971, p. 445): “The upshot of the problem is that theories rarely die in psychology, they just accumulate”.

Two causes for the lack of theory development are discussed in the first crisis. The first cause is an overreliance on statistical and methodological developments, and too little thought about the theoretical questions that are being asked. Cartwright (1979, p. 87) notes: “But these impressive gains in technical competence and sophistication have been, I fear, something of a mixed blessing, for the fascination with technique seems all too often to have replaced a concern for substantive significance. The literature is full of studies that do little more than demonstrate the technical virtuosity of the investigator, and one might think that our journals would have to go out of business if use of the analysis of variance were to be prohibited.”. Similarly, “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” (M. B. Smith, 1972, p. 94). This concern still plays a role

during the second crisis, as for example expressed by Gigerenzer (2018, p. 17): “Second, and most important, obtaining statistical significance has become a surrogate for good scientific practice, pushing principles such as formulating precise theories, conducting double-blind experiments, minimizing measurement error, and replicating findings into the sidelines.”

Psychology was seen as becoming an accumulation of facts, without integrating these facts into theories: “The respect of common sense, the proliferation of experimental studies lacking theoretical preoccupations, and the isolation of various areas of research in social psychology combine to explain the accumulation of facts and notions which do not amount to real progress since they are not conceptually integrated and since no theory is, in any real sense, disconfirmed or replaced by another” (Moscovici, 1972).

When it comes to the quality of theories in psychology, there seems to have been little progress. In a 2004 special issue that had the goal to strengthen training in theory construction in social and personality psychology, Kruglanski and Higgins (2004, p. 96) write “Theory construction plays an essential role in any field of science. We believe that social personality psychology has tended over time to be more phenomenon and data driven than theory driven and that this state of affairs may have impeded scientific progress in our field”. The continued criticism on theorizing in social psychology was repeated during the second crisis: “we argue that the crisis runs much deeper and is ultimately rooted in theory or lack thereof. Many subfields within psychology (though not all!) lack any overarching, integrative general theoretical framework that would allow researchers to derive specific predictions from more general premises” (Muthukrishna & Henrich, 2019, p. 221). Similarly, Oberauer and Lewandowsky (2019, p. 1596) write: “Here, we argue that, in addition to poor methods, the replication crisis is also due to the prevalence of theories that have only a weak logical relation to the hypotheses through which they are evaluated empirically. We suggest that the replication crisis is best resolved by focusing attention on the role of theorizing, and we do not

believe that current recommendations that focus entirely on data generation are sufficient to overcome the crisis.” In addition to a replication crisis, researchers speak of a theory crisis (Eronen & Bringmann, 2021).

Researchers discuss how the reward structures in psychology favor ‘fun’ studies over studies that make a theoretical contribution. The observations of Ring (1967, p. 117) during the first crisis are worth citing in full: “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. 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” and “The implicit values that produce this sort of research include the following: 1. Experiments should be as flashy and flamboyant as possible. 2. If you can think of an effective manipulation, fine; if you can think of an effective manipulation that is also amusing, even better. 3. If the topic selected for study is itself prosaic, you should reconsider; if you go ahead, at least study it cleverly. 4. Never make an obvious prediction.” Similarly, Pereboom (1971, p. 449) writes “In the light of all the previous discussion and the evidence, it no longer seems productive to continue to invent “cute” experiments or increase the sophistication of our experimental designs while decreasing the breadth of our theories”.

These concerns are echoed in the second crisis by Ledgerwood and Sherman (2012, p. 63): “Placing a premium on brief reports of flashy findings can exacerbate the issue of declining effects, as single-study findings that lack a theoretical context or that contradict previous data are especially likely to be spurious”. Similarly, Giner-Sorolla (2019, p. 14) explicitly references the work by Ring from the first crisis when stating “Social psychologists with very different opinions on the crisis of evidence have voiced concerns about publishing



attention-grabbing, media-friendly research with surprising or “sexy” results – recapitulating Ring’s(1967) concerns about “‘fun-and games’ social psychology” at the onset of the previous crisis.”

The fact that almost identical concerns about the state of psychological theorizing are raised in the first and the second crisis is a source of concern, especially given the lack of practical reform that aims to improve good theorizing. Where the replication crisis has led to notable changes in the number of replication studies that are performed and published, journal policies about publishing close replication studies, funding for replication research, reflections on which studies are worthwhile to replicate (Isager et al., 2021; Pittelkow et al., 2023), transparent reporting of statistical analyses, and discussions about how to resolve ambiguities about diverging results across close replication studies (Uygun Tunç & Tunç, 2022), there are as of yet no similar concrete improvements implemented that can be expected to improve theorizing in psychology. Concrete suggestions to improve theory development in psychology deserve much greater attention.

### **Is there a crisis of relevance?**

During the previous crisis some of the most eminent psychologists publicly doubted the societal impact psychologists had. In a presidential address to the American Psychological Association George Miller (1969, p. 1063) stated: “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.” In part, this criticism was a consequence of the increased reliance on lab-based experiments (as opposed to more naturalistic paradigms or field studies), as illustrated by Deutsch (1976, p. 136): “current theory is deficient in characterizing both the socially relevant properties of individual

personalities and the psychologically relevant attributes of social situations.” The criticism on the lack of societal impact that psychology had was surprisingly widespread during the first crisis. Helmreich (1975, p. 558) writes “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.”

Doing more applied research is proposed as a solution to the problems in psychology by Pereboom (1971, p. 451): “In applied research our objectives can often be made very explicit. This in itself solves many problems.”

At the same time, it is observed that psychology needs to improve to generate knowledge that will be applicable. Aron (1979, p. 50) notes “Even though we have discovered quite a bit in the limited areas we have been able to research, what we have found has not been successful in application. It is important to note, however, that this is not because the principles we found were wrong — when they have been applied they work. Rather, the difficulty has been that the kinds of laws we have found through our experiments, laws governing the social influences on individuals, either cannot or will not be applied by society.” One reason for the lack of more relevant research is dominance of “hit-and-run” studies (Kruglanski, 1975). In the provocatively titled article “Why social psychology fails” Silverman (1977, p. 355) notes: “First, we note that the experiment, defined in the classic sense of manipulation of the independent variable, remains the method of choice. This would not necessarily be disquieting, except as it relates to a second observation, that social psychologists in the field continue to work in the pattern of heterogeneous, short term studies, of very limited impact on subjects, that predominated in the laboratory.”

Psychologists continued to raise concerns about how lab-based studies remained too far removed from actual behavior to make psychology practically applicable (e.g., Baumeister et al., 2007), or how psychological findings are often not ready to inform policy (IJzerman et

al., 2020), although there have also been defenses of the practical relevance of psychological research (Lilienfeld, 2012; Zimbardo, 2004). Nevertheless, during the second crisis the relevance of the field for societal wellbeing was rarely at the center of attention. Two exceptions come from Giner-Sorolla (2019), who reviews concerns on the relevance of the questions psychologists ask, the generalizability of findings due to the samples that are collected, and the way psychological science is communicated to the public, and Berman and Wilson (2021, p. 4), who note: “Our casual observation is that psychological theory has become unmoored from the guiding principle of practicality and is drifting toward more nuanced or myopic theoretical questions that are less relevant to helping solve the problems that people care about”. During the second crisis, doing more applied research is seen as a possible solution to concerns about generalizability (Yarkoni, 2021).

It is surprising that the societal relevance of psychological science not a topic of discussion in the second crisis, as not much has changed with respect to the research that is done. In response to the concerns raised by Ring (1967) about the prevalence of flashy experiments, McGuire (1967, p. 126) predicted the following: “I foresee that the ingenuity now exercised in creating laboratory surrogates for the real world will be steadily replaced by equal ingenuity exerted in utilizing the natural environment as a field in which to test these deductions.” Early meta-scientific studies examining the prevalence of field studies showed there was no increase between 1961 and 1970 (Fried et al., 1973; Higbee & Wells, 1972). The situation has not noticeably improved in the years since, and the hope that psychologists would perform more field studies in the future has not materialized. In general, researchers today are still concerned about the pressure to publish small but flashy studies: “Currently, swiftly publishing several small but catchy contributions trumps publishing a carefully planned multi-study investigation, because institutional incentives require boosting one’s publication count to excel in short-sighted performance metrics.” (Rahal et al., 2023, p. 165).

Although published articles where researchers express their doubts about the relevance of psychological science are largely absent during the second crisis, I doubt these concerns have disappeared. Instead, given that this topic regularly comes up in conversations I have with fellow academics, the concern might still exist, but is no longer discussed openly. Perhaps we have become less comfortable expressing our doubts about the relevance of our field publicly, but the topic is important enough to attempt to bring it more in the open for those researchers who are concerned about the societal relevance of our field.

### **Is there a generalizability crisis?**

Related to concerns about the relevance of psychological research for society, researchers expressed a concern during the first crisis that psychological findings failed to generalize. Silverman (1971, p. 583) noted: “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.”. Similarly, Epstein (1980, p. 790) writes: “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.” One concern about the lack of generalizability of psychological science was a direct consequence of Gergen’s (1973) argument that psychology should be seen as history, as psychology deals with findings do not remain stable over time, and effects are not repeatable. It should be noted that this idea, although popular among some, was also strongly criticized at the time, with for example Schlenker (1974, p. 2) stating “Gergen's arguments typify several popular misconceptions about the nature of science and exhibit some undue pessimism”. Another concern is the lack of representativeness in participants (Carlson, 1971; Triandis, 1975). For example, McGuire

(1967, p. 313) writes: “Probably the most important of the difficulties is the growing realization among laboratory experimenters that our work is troubled by artifacts that make generalization and theoretical interpretation difficult.” He continues: “We have almost grown used to the embarrassment occasioned by the concentration of our research on college students. The proverbial “psychology of the college sophomore” is more worrisome as to generalizability for the social psychologist than for those working in perception, learning, etc.”. Schultz (1969) agrees: “In reading our journals, one receives the distinct impression that the only kind of people of interest to psychologists are college students! If college students were truly representative samples of the population at large, there would be no problem in generalizing from the results of our studies. But (fortunately or unfortunately) they do differ in highly significant ways from the general population, and we cannot have a truly meaningful science of human behavior by studying such a restricted sample.”

A meta-scientific study examining the common belief that field studies are more generalizable than lab experiments observed that both field studies and lab experiments sample from a limited range of subject, settings, and behaviors, and that both therefore provide little guarantees that observed findings will generalize. (Dipboye & Flanagan, 1979). Helmreich (1975, p. 551) raises the additional concern that it is difficult to generalize from the acute responses in the laboratory to effects on a longer timescale: “Subjects are usually held for at most one or two hours, leaving the researcher forced to generalize from a very short time frame to trends of interaction and response over months or years.” Some came to the defense of artificial lab-based experiments (a notable example is Mook’s (1983) ‘In defense of external invalidity’), but the lack of generalizability of research findings in psychology remained a concern, especially among those psychologists who also lamented the lack of practical relevance of the field.

Very little has been done to address these concerns in psychology, and the same concerns have been raised during the second crisis by Yarkoni (2021, p. 12): “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”.

### **Is there a methodological crisis?**

Concerns about methodological and statistical practices that inflate the probability of observing a false positive (or Type 1 error) are centuries old (Babbage, 1830). Neher (1967) coined the term ‘probability pyramiding’ for the inflated Type 1 error rate due to the practice of researchers to perform multiple tests, and the combined effects of Type 1 error inflation during the data analysis with the practice to selectively write up studies with significant findings, and publication bias favoring statistically significant results. Foreshadowing the well-known article “Why most published research findings are false” by Ioannidis (2005), Neher wrote: “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.”

In 1976 Barber published the book “Pitfalls in Human Research: Ten Pivotal Points”. In chapter 4 on “Investigator Data Analysis Effect” he presents nine potential issues: 1) performing unplanned analyses, 2) hypothesizing after results are known (cf. Kerr, 1998), 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) relatedly, not reporting non-significant results, 8) checking for errors after negative results, but not checking for errors after positive results, and 9) reporting statistically significant but practically insignificant effects. Several of these issues are remarkably similar to ‘researchers degrees of freedom’ presented in the seminal article on false positive psychology that played an important role in increasing awareness of research practices that inflate the probability of false positive results (Simmons et al., 2011), and the remaining issues are statistical shortcomings that remain a problem today.

Despite the fact that data analysis practices that inflate the Type 1 error rate were known, there was basically no discussion of these problematic research practices during the first crisis. Methodological criticism did play an important role in the first crisis (Faye, 2012; Higbee & Wells, 1972, 1972; McGuire, 1973), but the criticism was focused on how demand effects and experimenter effects introduced confounds in lab experiments (Orne, 1962; Rosenthal, 1966). The realization that thoughts and feelings of both participants and experimenters could introduce systematic bias was a real shock to social psychologists in the first crisis (Argyris, 1968; Schultz, 1969), and increased the criticism on lab-based studies. At the same time, the methodological crisis largely ignored research practices during the data analysis that inflate the Type 1 error rate, as discussed by Barber (1976). This is ironic, given that Barber & Silver (1968) raised the criticism that the evidence of experimenter effects itself suffered from inflated Type 1 error rates due to flexibility in the data analysis. We find a hint of a recognition of the problem in M. B Smith (1972, p. 94) when discussing how  $p$  values are not as tightly linked to support for our theories as we typically pretend, because “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”. But it is fair to say that these bad research practices were not publicly discussed. This is

regrettable, as an honest discussion of these problematic practices during the first crisis could have prevented considerable research waste.

The second crisis started with the realization that commonly used research practices allowed researchers to make scientific claims that turned out to be less reliable than researchers realized. The publication of an article describing 9 studies that provided empirical support for extra-sensory perception, (Bem, 2011), as well as the creation of open access journals that published failed replication studies of prominent findings in the field (Doyen et al., 2012) started a widespread discussion about problematic methodological practices (LeBel & Peters, 2011; Wagenmakers et al., 2011). Accessible articles that explained how problematic research practices increased the Type 1 error rate (Simmons et al., 2011), together with surveys suggesting many researchers admitted to using these practices (John et al., 2012) caused at least as great a shock among psychologists as the realization that participants and experimenters had thoughts and feelings had in the first crisis. The fact that these issues seem obvious in hindsight, and yet persisted for decades, is a good reason to display intellectual humility. After all, who knows which practice that we currently all engage in will be at the root of the next crisis in psychology?

### **Is the crisis caused by bad incentives?**

The belief that the incentive system in academia puts pressure on researchers to prioritize quantity over quality is at least a century old. Case (1927, p. 325) wrote: “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!”.” Elms discusses pressures from outside academia to be relevant, the realization of their own biases due to the emancipation of women and racial minorities, and concerns about ethical standards in the field as a cause of the feeling of a crisis. He remarks (1975, p. 972): “ “The



contributions other factors have made to the crisis of confidence are less clear: for instance, the publish-or-perish pressures and shrinking job market within the academic world”. Still, he also believes that “ social psychologists must find ways of adjusting to or moderating academic pressures to publish, at the same time that they reduce their own rate of publishing findings that are so incomplete as to be misleading.” Sherif is clearer on the negative effect of the incentive system on the quality of research (1977, p. 371): “Even from this quick glance at the state of two major areas of research, a rather clear but somewhat disturbing picture stands out; piles and piles of output in research and publication, accelerated almost daily by the ‘publish or perish’ policy of academia. In brief, the ratio of golden kernels to be saved to the huge quantity of chaff is only a minute fraction.” Similarly, Mahoney (1979, p. 357) notes: “Of course, this tendency to publish (and speculate) quickly is reinforced by the reward system in science, which gives strong encouragement to volume, as well as to velocity. Studies of the academic scientist leave little doubt that ‘publish or perish’ is more than a glib alliteration.”

The ‘publish or perish’ culture is often mentioned in articles written during the second crisis (e.g., Chambers, 2019; Fraser et al., 2018; Rahal et al., 2023; Sarafoglou et al., 2022; Schimmack, 2012). What’s new is that the incentive structure is not just presented as a cause of some of the problems that the discipline faces (Fiedler et al., 2012; Giner-Sorolla, 2019; Heesen, 2018; Koole & Lakens, 2012; Romero, 2019), but as a malleable aspect of the scientific ecosystem that can be changed (e.g., Maner, 2014; Nosek et al., 2012; Schimmack, 2012; Tackett et al., 2017; Tiokhin et al., 2023). Meta-scientists start to experiment with a range of possible interventions to reward good research practices, such as the Registered Report publication format (Chambers et al., 2014; Nosek & Lakens, 2014), badges that identify articles that are preregistered or have open data (Kidwell et al., 2016), and peer reviewers who commit to only review scientific articles that either share data or explain why

data can not be shared (Morey et al., 2016). Perhaps due to the greater number and diversity of scientific journals, there is a lot more experimentation in journal policies and publication formats. It is often too early to be able to empirically evaluate the consequences of these interventions, but initial meta-scientific research suggests some novel developments have achieved their aim. For example, Registered Reports lead to more non-significant results being published, while the articles are at least as high-quality as standard publications (Allen & Mehler, 2019; Scheel et al., 2021; Soderberg et al., 2021).

One bad incentive that deserved particular attention is the tendency of psychological scientists to prefer positive results over negative results, and publish the former, and fail to publish the latter. In a section 'The Secrets We Keep' Dunnette (1966, p. 347) writes: "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." Greenwald (1975) performed a meta-scientific study surveying reviewers and authors of articles submitted to the *Journal of Personality and Social Psychology* to provide the first model and parameter estimates of the causes of publication bias.

Publication bias, especially together with Type 1 error rate inflation and the choice to study unlikely hypotheses, can lead to a low probability that published findings are true effects (Ioannidis, 2005; Neher, 1967; Wacholder et al., 2004). This realization led to the increase of the application of statistical methods to identify bias in the scientific literature (Francis, 2014; Schimmack, 2012; Simonsohn et al., 2014). For example, Carter and McCullough (2014) re-analyzed a published meta-analysis of 198 independent tests of the ego-depletion hypothesis and identified strong bias in the studies included in the meta-

analysis. Estimates of the unbiased effect size estimate raised the possibility there was no effect at all. In 2016 a preregistered replication study with 2141 participants found a non-significant ego-depletion effect that was basically zero (Hagger et al., 2016). In 2021 a preregistered replication study with 3531 participants, performed by original authors of ego-depletion studies, also found a non-significant effect very close to zero (Vohs et al., 2021). This research line empirically demonstrated the extreme consequences publication bias can have on the scientific record, where there can be 198 studies indicating there is an effect, with unbiased replication studies suggesting there was no effect all along. Such findings strongly contributed to the sense of a replication crisis, as researchers were increasingly uncertain about whether the research they built on would also turn out to be irreplicable.

## **Discussion**

After two crises, what has changed? With respect to methodological practices the first crisis made psychologists more acutely aware of demand effects and experimenter effects. At the same time, with the sharp increase of computer-based experimental paradigms, and a move away from extensive deception-based experiments, the influence of experimenter effects is less of an issue than it was during the first crisis. The second crisis created widespread awareness of research practices that lead to unreliable statistical inferences, the importance of Type 1 and Type 2 error control, and the detrimental effects of publication bias. With respect to generalizability the first crisis led to the social constructionist movement (Gergen, 1985). Even if it might not have become the dominant approach in psychology (Stroebe & Kruglanski, 1989), it forced mainstream social psychologists to acknowledge the inherent context-dependency of their field, up to the point that if you can today summarize all knowledge generated by social psychology in two words: It depends. At the same time, there is room for improvement as researchers still tend to overgeneralize in the claims they make (Simons et al., 2017; Yarkoni, 2021). Further attention is being given to how we can design

studies that allow for more generalizable conclusions (Almaatouq et al., 2022; Baribault et al., 2018).

The second crisis led to a burgeoning field of meta-science. Faust (2005) credits Paul Meehl and himself for the invention of meta-science, and defines the field as follows: “Meta-science, in essence, involves the application of more rigorous methods than are presently used in the history and philosophy of science to answer questions about how science does and does not work best”. Of course, meta-science has a much longer history, but it has seen an immense increase in popularity since the second crisis, establishing itself as its own field, with dedicated scientific journals (Carlsson et al., 2017; Simons, 2018). Both crises have provided fertile ground for new scientific approaches and research areas.

Doubts about the relevance of psychological science have largely disappeared from public discourse, and do not seem to be a central concern in the second crisis in psychology. The idea that the incentives will need to change to increase the relevance of psychological science (Berkman & Wilson, 2021; Giner-Sorolla, 2019) has gone unchallenged. The recent explosion of research into how social and behavioral science could contribute to the COVID-19 response provided a good test case of the relevance of psychological research for society (Bavel et al., 2020). Concerns about the lack of generalizability of psychological findings were raised (Bryan et al., 2021; IJzerman et al., 2020), and mirror the concerns about generalizability that were so important in the first crisis. Despite thousands of studies, and the obvious potential of psychological knowledge to change behavior during a pandemic, I am not aware of any success stories where psychological knowledge made a substantial contribution to the way the world dealt with COVID-19. This makes the absence of a discussion about the relevance of psychological science during the second crisis even more surprising. Perhaps there is a new generation of psychological scientists who value practical relevance less compared to the psychologists active during the first crisis. It is also possible

that concerns about the relevance of psychological science are still present, but are simply not publicly discussed. This topic deserves further attention, as it also remained unresolved during the first crisis. Just as how the absence of an honest discussion about inflated Type 1 error rates during the first crisis came back to haunt us in the second crisis, lack of an honest discussion concerning the societal relevance of our field during the second crisis might come back to haunt us in a third crisis.

With respect to theory there has arguably been little progress, despite repeated reminders that psychologists should improve their theorizing. There have been no notable changes in journal policies to promote better theorizing during the second crisis, and there is a lack of meta-scientific work on what makes theories in the literature more successful, and what makes them less successful. Meehl's work on cliometric metatheory (Meehl, 1992, 2002, 2004) offers an exciting and as of yet unexplored approach for meta-scientists interested in improving theorizing in psychological science. The theory crisis is not the only fundamental issue that has remained unresolved after both crises, as there are at least equally fundamental issues concerning measurement and validity that have received very little attention in the first crisis, and far too little attention in the second crisis (Flake et al., 2017; Schimmack, 2021; Vazire et al., 2022). I would not be surprised if issues related to measurement and validity have to wait until the third crisis in psychology before they receive the attention they deserve.

With respect to concerns about the incentive structure in academia only minor changes were introduced during the first crisis. Examples are requiring authors to report methodological procedures in more detail to enable direct replications (Greenwald, 1976), and the launch of the Journal of Applied Social Psychology to stimulate research with more direct relevance for society (Helmreich, 1975). Much more substantial changes to the reward structure have been introduced during the second crisis, although it is too early to tell if these

will prove effective. The second crisis has coincided with rapid technological advances that have made greater transparency possible, and has led to the emergence of the Open Science movement. The internet has also facilitated team science, which has enabled large scale replication project (Klein et al., 2014; Open Science Collaboration, 2015), more representative data collection from laboratories across the world with the goal to improve generalizability (Moshontz et al., 2018), and adversarial collaborations in which theoretical predictions are sharpened and tested (Coles et al., 2020). These developments are still the exception rather than the rule. The future will tell if these changes to the way psychologists work will become more popular over time, and realize their potential to improve psychological science.

While for some researchers the fundamental problems that were discussed during the first and second crisis deserve all the attention they have received, others are concerned and annoyed when their discipline is faced with a widespread crisis narrative. In a footnote, Zajonc reflects on the first crisis (1989, p. 347): “Some expressions of epistemological malaise have been outright destructive and paralysing for social psychology. Entirely without heuristic consequence has been the argument about ‘social psychology as history’ [...]. Nevertheless, the controversy has discouraged promising students from entering the field and granting agencies from increasing social psychological research budgets.” Similar concerns for negative effects of a crisis narrative have been expressed during the second crisis by Baumeister (2016, p. 156): “Another set of potential winners are researchers in nearby fields who compete with social psychologists for grant funds, awards, faculty lines, graduate student support, and pages in broad-readership journals. As social psychologists persist in discrediting our field's work, we lend ammunition to other areas who argue that the precious and scarce research funds should be diverted away from social psychology and into their own areas.”

Although these concerns are understandable, sometimes events happen that fundamentally challenge a scientific discipline, and the concerns that emerge are too widespread to prevent a temporary negative effect on the reputation of a discipline. Of course, the incentive structure equally exerts its pressure on those who claim there is a crisis, as Sherif (1977, p. 372) already observed: “There is the danger of further accumulation of a whole big crop of crisis literature because it is attention getting; it is likely to insure its author handsome professional mileage.” Yet, such skepticism seems unwarranted. There has been too much agreement throughout and across the two crises about the ways in which psychological science can be improved to dismiss these concerns as careerism. In a conservative system such as science it might require a crisis to instigate self-reflection on how a field can improve (McGuire, 1973; Mills, 1979), and to motivate scientists to embrace change.

## References

- Ahlgren, A. (1969). A modest proposal for encouraging replication. *American Psychologist*, 24(4), 471. <https://doi.org/10.1037/h0037798>
- Aldhous, P. (2011). *Journal rejects studies contradicting precognition*. New Scientist. <https://www.newscientist.com/article/dn20447-journal-rejects-studies-contradicting-precognition/>
- Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, 17(5), e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Almaatouq, A., Griffiths, T. L., Suchow, J., Whiting, M. E., Evans, J., & Watts, D. J. (2022). *Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences*. PsyArXiv. <https://doi.org/10.31234/osf.io/anjkm>
- Argyris, C. (1968). Some unintended consequences of rigorous research. *Psychological Bulletin*, 70(3, Pt.1), 185–197. <https://doi.org/10.1037/h0026145>
- Aron, A. (1979). How social psychology can succeed: A response to Silverman. *Canadian Psychological Review/Psychologie Canadienne*, 20(1), 48–52. <https://doi.org/10.1037/h0081492>
- Babbage, C. (1830). *Reflections on the Decline of Science in England: And on Some of Its Causes*. B. Fellowes.
- Bakan, D. (1967). *On method: Toward a reconstruction of psychological investigation*. San Francisco, Jossey-Bass. [http://archive.org/details/onmethodtowardre0000baka\\_y0b7](http://archive.org/details/onmethodtowardre0000baka_y0b7)
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), Article 7604. <https://doi.org/10.1038/533452a>
- Barber, T. X. (1976). *Pitfalls in Human Research: Ten Pivotal Points*. Pergamon Press.



- Barber, T. X., & Silver, M. J. (1968). Fact, fiction, and the experimenter bias effect. *Psychological Bulletin*, 70(6, Pt.2), 1–29. <https://doi.org/10.1037/h0026724>
- Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., Ravenzwaaij, D. van, White, C. N., Boeck, P. D., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, 115(11), 2607–2612. <https://doi.org/10.1073/pnas.1708285114>
- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153–158. <https://doi.org/10.1016/j.jesp.2016.02.003>
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*, 2(4), 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., & Druckman, J. N. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460–471.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425. <https://doi.org/10.1037/a0021524>
- Berkman, E. T., & Wilson, S. M. (2021). So useful as a good theory? The practicality crisis in (social) psychological theory. *Perspectives on Psychological Science*, 1745691620969650. <https://doi.org/10.1177/1745691620969650>

- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 1–10.  
<https://doi.org/10.1038/s41562-021-01143-3>
- Carlson, R. (1971). Where is the person in personality research? *Psychological Bulletin*, 75(3), 203–219. <https://doi.org/10.1037/h0030469>
- Carlsson, R., Danielsson, H., Heene, M., Innes-Ker, A., Lakens, D., Schimmack, U., Schönbrodt, F. D., van Assen, M., & Weinstein, Y. (2017). Inaugural Editorial of Meta-Psychology. *Meta-Psychology*, 1, 1–3. <https://doi.org/10.15626/MP2017.1001>
- Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00823>
- Cartwright, D. (1979). Contemporary Social Psychology in Historical Perspective. *Social Psychology Quarterly*, 42(1), 82–93. <https://doi.org/10.2307/3033880>
- Case, C. M. (1927). Scholarship in sociology. *Sociology and Social Research*, 12, 323–340.
- Chambers, C. D. (2019). The registered reports revolution Lessons in cultural reform. *Significance*, 16(4), 23–27. <https://doi.org/10.1111/j.1740-9713.2019.01299.x>
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1(1), 4–17.  
<https://doi.org/10.3934/Neuroscience.2014.1.4>
- Chartier, C. R., Riegelman, A., & McCarthy, R. J. (2018). StudySwap: A Platform for Interlab Replication, Collaboration, and Resource Exchange. *Advances in Methods and Practices in Psychological Science*, 1(4), 574–579.  
<https://doi.org/10.1177/2515245918808767>

- Coles, N. A., March, D. S., Marmolejo-Ramos, F., Banaruee, H., Butcher, N., Cavallet, M., Dagaev, N., Eaves, D., Foroni, F., Gorbunova, E., Gygax, P., Poveda, J. A. H., Ikeda, A. A., Kathin-Zadeh, O., Özdoğru, A. A., Parzuchowski, M., Ruiz-Fernández, S., Som, B., Suarez, I., ... Marozzi, M. (2020). The Many Smiles collaboration: A multi-Lab foundational test of the facial feedback hypothesis. *Nature Human Behaviour*.  
<https://doi.org/10.31234/osf.io/cvpuw>
- Crandall, C., & Sherman, J. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66.  
<https://doi.org/10.1016/j.jesp.2015.10.002>
- Deutsch, M. (1976). Theorizing in Social Psychology. *Personality and Social Psychology Bulletin*, 2(2), 134–141. <https://doi.org/10.1177/014616727600200214>
- Dipboye, R. L., & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist*, 34(2), 141.
- Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral Priming: It's All in the Mind, but Whose Mind? *PLOS ONE*, 7(1), e29081.  
<https://doi.org/10.1371/journal.pone.0029081>
- Dunnette, M. D. (1966). Fads, fashions, and folderol in psychology. *American Psychologist*, 21(4), 343–352. <https://doi.org/10.1037/h0023535>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621.  
<https://doi.org/10.3389/fpsyg.2015.00621>
- Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D.-J., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, L. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrichetto, L., Arnal, J. D.,

- Arrow, H., Babincak, P., ... Nosek, B. A. (2020). Many Labs 5: Testing Pre-Data-Collection Peer Review as an Intervention to Increase Replicability. *Advances in Methods and Practices in Psychological Science*.  
<https://doi.org/10.1177/2515245920958687>
- Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, 30(10), 967.
- Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist*, 35(9), 790–806. <https://doi.org/10.1037/0003-066X.35.9.790>
- Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, 1745691620970586.  
<https://doi.org/10.1177/1745691620970586>
- Faust, D. (2005). Why Paul Meehl will revolutionize the philosophy of science and why it should matter to psychologists. *Journal of Clinical Psychology*, 61(10), 1355–1366.  
<https://doi.org/10.1002/jclp.20185>
- Faye, C. (2012). American social psychology: Examining the contours of the 1970s crisis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(2), 514–521.  
<https://doi.org/10.1016/j.shpsc.2011.11.010>
- Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The Long Way From  $\alpha$ -Error Control to Validity Proper: Problems With a Short-Sighted False-Positive Debate. *Perspectives on Psychological Science*, 7(6), 661–669. <https://doi.org/10.1177/1745691612462587>
- Fishman, D. B., & Neigher, W. D. (1982). American psychology in the eighties: Who will buy? *American Psychologist*, 37(5), 533–546. <https://doi.org/10.1037/0003-066X.37.5.533>

- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>
- Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, 29(2), 158–181. <https://doi.org/10.1177/0959354319835322>
- Francis, G. (2014). The frequency of excess success for articles in Psychological Science. *Psychonomic Bulletin & Review*, 21(5), 1180–1187. <https://doi.org/10.3758/s13423-014-0601-x>
- Frank, M. C., & Saxe, R. (2012). Teaching Replication. *Perspectives on Psychological Science*, 7(6), 600–604. <https://doi.org/10.1177/1745691612460686>
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLOS ONE*, 13(7), e0200303. <https://doi.org/10.1371/journal.pone.0200303>
- Fried, S. B., Gumpfer, D. C., & Allen, J. C. (1973). Ten years of social psychology: Is there a growing commitment to field research? *American Psychologist*, 28(2), 155–156. <https://doi.org/10.1037/h0034202>
- Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26, 309–320. <https://doi.org/10.1037/h0034436>
- Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40(3), 266–275. <https://doi.org/10.1037/0003-066X.40.3.266>
- Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, 1(2), 198–218. <https://doi.org/10.1177/2515245918771329>

- Giner-Sorolla, R. (2012). Will We March to Utopia, or Be Dragged There? Past Failures and Future Hopes for Publishing Our Science. *Psychological Inquiry*, 23(3), 263–266.  
<https://doi.org/10.1080/1047840X.2012.706506>
- Giner-Sorolla, R. (2019). From crisis of evidence to a “crisis” of relevance? Incentive-based answers for social psychology’s perennial relevance worries. *European Review of Social Psychology*, 30(1), 1–38. <https://doi.org/10.1080/10463283.2018.1542902>
- Goertzen, J. R. (2008). On the Possibility of Unification: The Reality and Nature of the Crisis in Psychology. *Theory & Psychology*, 18(6), 829–852.  
<https://doi.org/10.1177/0959354308097260>
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82(1), 1–20. <https://doi.org/10.1037/h0076157>
- Greenwald, A. G. (Ed.). (1976). An editorial. *Journal of Personality and Social Psychology*, 33(1), 1–7. <https://doi.org/10.1037/h0078635>
- Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., ... Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*, 11(4), 546–573.  
<https://doi.org/10.1177/1745691616652873>
- Hales, A. H. (2016). Does the conclusion follow from the evidence? Recommendations for improving research. *Journal of Experimental Social Psychology*, 66, 39–46.  
<https://doi.org/10.1016/j.jesp.2015.09.011>
- Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, 42(5), 443–455.  
<https://doi.org/10.1037/0003-066X.42.5.443>

- Heesen, R. (2018). Why the Reward Structure of Science Makes Reproducibility Problems Inevitable. *Journal of Philosophy*, 115(12), 661–674.  
<https://doi.org/10.5840/jphil20181151239>
- Helmreich, R. (1975). Applied Social Psychology: The Unfulfilled Promise. *Personality and Social Psychology Bulletin*, 1(4), 548–560.  
<https://doi.org/10.1177/014616727500100402>
- Higbee, K. L., & Wells, M. G. (1972). Some research trends in social psychology during the 1960s. *American Psychologist*, 27(10), 963–966. <https://doi.org/10.1037/h0033453>
- Hunt, K. (1975). Do we really need more replications? *Psychological Reports*, 36(2), 587–593.
- IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11), Article 11. <https://doi.org/10.1038/s41562-020-00990-w>
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Isager, P. M., Aert, R. C. M. van, Bahník, Š., Brandt, M., DeSoto, K. A., Giner-Sorolla, R., Krueger, J., Perugini, M., Ropovik, I., Veer, A. van 't, Vranka, M. A., & Lakens, D. (2021). Deciding what to replicate: A formal definition of “replication value” and a decision model for replication study selection. *Psychological Methods*.  
<https://doi.org/10.31222/osf.io/2gurz>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>

- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217.  
[https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology*, 14(5), e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., ... Nosek, B. A. (2014). Investigating Variation in Replicability: A “Many Labs” Replication Project. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Koole, S. L., & Lakens, D. (2012). Rewarding replications A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7(6), 608–614.  
<https://doi.org/10.1177/1745691612462586>
- Kruglanski, A. W. (1975). Theory, experiment and the shifting publication scene in personality and social psychology. *Personality and Social Psychology Bulletin*, 1(3), 489–492. <https://doi.org/10.1177/014616727500100305>
- Kruglanski, A. W., & Higgins, E. T. (2004). Theory Construction in Social Personality Psychology: Personal Experiences and Lessons Learned. *Personality and Social Psychology Review*, 8(2), 96–97. [https://doi.org/10.1207/s15327957pspr0802\\_1](https://doi.org/10.1207/s15327957pspr0802_1)



- LeBel, E. P., Berger, D., Campbell, L., & Loving, T. J. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology*, 113, 254–261.  
<https://doi.org/10.1037/pspi0000106>
- LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15(4), 371–379. <https://doi.org/10.1037/a0025172>
- Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science*, 7(1), 60–66.
- Lewin, M. A. (1977). Kurt Lewin's View of Social Psychology: The Crisis of 1977 and the Crisis of 1927. *Personality and Social Psychology Bulletin*, 3(2), 159–172.  
<https://doi.org/10.1177/014616727700300203>
- Lilienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist*, 67(2), 111–129.  
<https://doi.org/10.1037/a0023963>
- Loevinger, J. (1968). The “information explosion.” *American Psychologist*, 23(6), 455–455.  
<https://doi.org/10.1037/h0020800>
- Lubin, A. (1957). Replicability as a publication criterion. *American Psychologist*, 12, 519–520. <https://doi.org/10.1037/h0039746>
- Luttrell, A., Petty, R. E., & Xu, M. (2017). Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology*, 69, 178–183. <https://doi.org/10.1016/j.jesp.2016.09.006>
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70(3, Pt.1), 151–159. <https://doi.org/10.1037/h0026141>

- Mahoney, M. J. (1979). Review Paper: Psychology of the Scientist: An Evaluative Review. *Social Studies of Science*, 9(3), 349–375.  
<https://doi.org/10.1177/030631277900900304>
- Maner, J. K. (2014). Let's Put Our Money Where Our Mouth Is: If Authors Are to Change Their Ways, Reviewers (and Editors) Must Change With Them. *Perspectives on Psychological Science*, 9(3), 343–351. <https://doi.org/10.1177/1745691614528215>
- Manis, M. (1975). Comment on Gergen's "Social Psychology as History". *Personality and Social Psychology Bulletin*, 1(2), 450–455.
- McGuire, W. J. (1967). Some impending reorientations in social psychology: Some thoughts provoked by Kenneth Ring. *Journal of Experimental Social Psychology*, 3(2), 124–139. [https://doi.org/10.1016/0022-1031\(67\)90017-0](https://doi.org/10.1016/0022-1031(67)90017-0)
- McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, 26(3), 446–456.  
<https://doi.org/10.1037/h0034345>
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- Meehl, P. E. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports*, 71, 339–339.
- Meehl, P. E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. *Psychological Reports*, 91(2), 339–404.  
<https://doi.org/10.2466/pr0.2002.91.2.339>
- Meehl, P. E. (2004). Cliometric metatheory III: Peircean consensus, verisimilitude and asymptotic method. *The British Journal for the Philosophy of Science*, 55(4), 615–643.

- Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist*, 24(12), 1063–1075. <https://doi.org/10.1037/h0028988>
- Mills, T. M. (1979). Changing paradigms for studying human groups. *The Journal of Applied Behavioral Science*, 15(3), 407–423.
- Minton, H. L. (1984). J. F. Brown's Social Psychology of the 1930s: A Historical Antecedent to the Contemporary Crisis in Social Psychology. *Personality and Social Psychology Bulletin*, 10(1), 31–42. <https://doi.org/10.1177/0146167284101003>
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, 38(4), 379–387.
- Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C. C., Newman, D. P., Schönbrodt, F. D., & others. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, 3(1), 150547.
- Moscovici, S. (1972). Society and theory in social psychology. In *Context of social psychology* (pp. 17–81).
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., & Antfolk, J. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 1(4), 501–515.
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3(3), 221–229.
- Nederhof, A. J., & Zwier, A. G. (1983). The 'crisis' in social psychology, an empirical approach. *European Journal of Social Psychology*, 13(3), 255–280. <https://doi.org/10.1002/ejsp.2420130305>

- Neher, A. (1967). Probability Pyramiding, Research Error and the Need for Independent Replication. *The Psychological Record*, 17(2), 257–262.  
<https://doi.org/10.1007/BF03393713>
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618. <https://doi.org/10.3758/s13423-019-01645-2>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776.
- Pashler, H., & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/1745691612465253>
- Pereboom, A. C. (1971). Some Fundamental Problems in Experimental Psychology: An Overview. *Psychological Reports*, 28(2). <https://doi.org/10.2466/pr0.1971.28.2.439>
- Pettigrew, T. F. (2018). The Emergence of Contextual Social Psychology. *Personality and Social Psychology Bulletin*, 44(7), 963–971.  
<https://doi.org/10.1177/0146167218756033>

- Pittelkow, M.-M., Field, S. M., Isager, P. M., van't Veer, A. E., Anderson, T., Cole, S. N., Dominik, T., Giner-Sorolla, R., Gok, S., Heyman, T., Jekel, M., Luke, T. J., Mitchell, D. B., Peels, R., Pendrous, R., Sarrazin, S., Schauer, J. M., Specker, E., Tran, U. S., ... van Ravenzwaaij, D. (2023). The process of replication target selection in psychology: What to consider? *Royal Society Open Science*, *10*(2), 210586.  
<https://doi.org/10.1098/rsos.210586>
- Pittelkow, M.-M., Hoekstra, R., Karsten, J., & van Ravenzwaaij, D. (2021). Replication target selection in clinical psychology: A Bayesian and qualitative reevaluation. *Clinical Psychology: Science and Practice*, *28*(2), 210–221.  
<https://doi.org/10.1037/cps0000013>
- Rahal, R.-M., Fiedler, S., Adetula, A., Berntsson, R. P.-A., Dirnagl, U., Feld, G. B., Fiebach, C. J., Himi, S. A., Horner, A. J., Lonsdorf, T. B., Schönbrodt, F., Silan, M. A. A., Wenzler, M., & Azevedo, F. (2023). Quality research needs good working conditions. *Nature Human Behaviour*, *7*(2), Article 2. <https://doi.org/10.1038/s41562-022-01508-2>
- Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*.
- Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect. *PLoS ONE*, *7*(3), e33423. <https://doi.org/10.1371/journal.pone.0033423>
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, *14*(11), e12633.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. Appleton-Century-Crofts.

- Samelson, F. (1980). J B Watson's Little Albert, Cyril Burt's twins, and the need for a critical science. *American Psychologist*, 35(7), 619–625. <https://doi.org/10.1037/0003-066X.35.7.619>
- Sarafoglou, A., Kovacs, M., Bakos, B., Wagenmakers, E.-J., & Aczel, B. (2022). A survey on how preregistration affects the research workflow: Better science but more work. *Royal Society Open Science*, 9(7), 211997. <https://doi.org/10.1098/rsos.211997>
- Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, 4(2), 25152459211007468. <https://doi.org/10.1177/25152459211007467>
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4), 551–566. <https://doi.org/10.1037/a0029487>
- Schimmack, U. (2021). The Validation Crisis in Psychology. *Meta-Psychology*, 5. <https://doi.org/10.15626/MP.2019.1645>
- Schlenker, B. R. (1974). Social psychology and science. *Journal of Personality and Social Psychology*, 29(1), 1–15. <https://doi.org/10.1037/h0035668>
- Schlosberg, H. (1951). *Repeating fundamental experiments*.
- Schultz, D. P. (1969). The human subject in psychological research. *Psychological Bulletin*, 72(3), 214–228. <https://doi.org/10.1037/h0027880>
- Senn, D. J. (1988). Myopic social psychology: An overemphasis on individualistic explanations of social behavior. *Journal of Social Behavior & Personality*, 3(2), 45–52.

- Sherif, M. (1977). Crisis in Social Psychology: Some Remarks Towards Breaking Through the Crisis. *Personality and Social Psychology Bulletin*, 3(3), 368–382.  
<https://doi.org/10.1177/014616727700300305>
- Silverman, I. (1971). Crisis in social psychology: The relevance of relevance. *American Psychologist*, 26(6), 583.
- Silverman, I. (1977). Why social psychology fails. *Canadian Psychological Review / Psychologie Canadienne*, 18(4), 353–358. <https://doi.org/10.1037/h0081451>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366.  
<https://doi.org/10.1177/0956797611417632>
- Simons, D. J. (2018). Introducing Advances in Methods and Practices in Psychological Science. *Advances in Methods and Practices in Psychological Science*, 1(1), 3–6.  
<https://doi.org/10.1177/2515245918757424>
- Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science*, 9(5), 552–555. <https://doi.org/10.1177/1745691614543974>
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*, 12(6), 1123–1128. <https://doi.org/10.1177/1745691617708630>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547.  
<https://doi.org/10.1037/a0033242>
- Smith, M. B. (1972). Is experimental social psychology advancing? *Journal of Experimental Social Psychology*, 8(1), 86–96.

- Smith, N. C. (1970). Replication studies: A neglected aspect of psychological research. *American Psychologist*, 25(10), 970–975. <https://doi.org/10.1037/h0029774>
- Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, 5(8), Article 8. <https://doi.org/10.1038/s41562-021-01142-4>
- Stroebe, W., & Kruglanski, A. W. (1989). Social psychology at Epistemological cross-roads: On Gergen's choice. *European Journal of Social Psychology*, 19(5), 485–489. <https://doi.org/10.1002/ejsp.2420190514>
- Stroebe, W., & Strack, F. (2014). The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science*, 9(1), 59–71. <https://doi.org/10.1177/1745691613514450>
- Sturm, T., & Mülberger, A. (2012). Crisis discussions in psychology—New historical and philosophical perspectives. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(2), 425–433. <https://doi.org/10.1016/j.shpsc.2011.11.001>
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., Oltmanns, T. F., & Shrout, P. E. (2017). It's Time to Broaden the Replicability Conversation: Thoughts for and From Clinical Psychological Science. *Perspectives on Psychological Science*, 12(5), 742–756. <https://doi.org/10.1177/1745691617690042>
- Tiokhin, L., Lakens, D., Smaldino, P. E., & Panchanathan, K. (2023). Shifting the Level of Selection in Science. In *Perspectives on Psychological Science*. <https://doi.org/10.31222/osf.io/juwck>
- Triandis, H. C. (1975). Social psychology and cultural analysis. *Journal for the Theory of Social Behaviour*, 5(1), 81–106. <https://doi.org/10.1111/j.1468-5914.1975.tb00344.x>



- Uygun Tunç, D., & Tunç, M. N. (2022). A Falsificationist Treatment of Auxiliary Hypotheses in Social and Behavioral Sciences: Systematic Replications Framework. In *Meta-Psychology*. <https://doi.org/10.31234/osf.io/pdm7y>
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6454–6459. <https://doi.org/10.1073/pnas.1521897113>
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility Beyond Replicability: Improving the Four Validities in Psychological Science. *Current Directions in Psychological Science*, 31(2), 162–168. <https://doi.org/10.1177/09637214211067779>
- Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Bunyi, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., ... Albarracín, D. (2021). A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect. *Psychological Science*, 32(10), 1566–1581. <https://doi.org/10.1177/0956797621989733>
- Wacholder, S., Chanock, S., Garcia-Closas, M., El ghormli, L., & Rothman, N. (2004). Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *JNCI Journal of the National Cancer Institute*, 96(6), 434–442. <https://doi.org/10.1093/jnci/djh075>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>

- Yarkoni, T. (2021). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37.  
<https://doi.org/10.1017/S0140525X20001685>
- Zajonc, R. B. (1989). Styles of explanation in social psychology. *European Journal of Social Psychology*, 19(5), 345–368. <https://doi.org/10.1002/ejsp.2420190503>
- Zimbardo, P. G. (2004). Does Psychology make a significant difference in our lives?  
*American Psychologist*, 59(5), 339–351. <https://doi.org/10.1037/0003-066X.59.5.339>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41.  
<https://doi.org/10.1017/S0140525X17001972>