

# Two Questions to Foster Critical Thinking in the Field of Psychology: Are There Any Reasons to Expect a Different Outcome, and What Are the Consequences If We Don't Find What We Were Looking For?

Peter Holtz

Leibniz-Institut für Wissensmedien IWM, Tübingen, Germany

There are many factors that contribute to the present crisis of confidence in psychology, among them epistemological causes: Under pressure to 'publish or perish' and to 'get visible or vanish' in order to survive in an increasingly globalized academic job market, psychologists may often be too eager to find their hypotheses confirmed by empirical data. They may also not pay enough attention to alternative theories and consequently often miss opportunities to learn from their failures to obtain the expected results in their studies. In this paper, I propose to start asking two questions physicist John Platt had proposed in 1964 on a regular basis in the field of psychology as a means of fostering critical thinking or to encourage a critical approach to the growth of scientific knowledge: Are there reasons to expect a different outcome, and what consequence is it going to have if the study does not yield the expected results? I explore what potential these two questions have for ensuring epistemological progress by asking them with respect to social-priming research, which is one of the research programmes that have recently been criticized in the course of the 'reproducibility debate'.

**Keywords:** Epistemology, philosophy of science, critical rationalism, falsificationism, strong inference, social priming, replicability, reproducibility, crisis of confidence.

In this paper I will argue that the reasons for the present 'crisis' in psychology can be attributed in part to epistemological deficiencies: Psychologists are too eager to find their theories corroborated by empirical evidence, they do not consider competing theories often enough, and they often do not pay enough attention to the inferences that can be drawn from not finding the expected results. However, scores of philosophers (e.g., Dewey, 1903/2004; Popper, 1934/1959) and scientists (e.g. Feynman, 1974) have argued that as scientists, and as human beings in general, we can learn most of all from our mistakes. As a remedy to psychologists' apparently overly optimistic approach to scientific research, I will introduce two questions that physicist John Platt (1964) once proposed as a means of accelerating progress in science. The questions are related to possible alternative theories and provoke

the researcher to consider empirical outcomes that are contrary to expectations. I will argue that these two questions can and should be asked with regard to any empirical study, and that the field of psychology would benefit from asking them on a regular basis. If researchers were to do so, critical thinking or a "critical approach" (Popper, 1962, p. 51) to the growth of knowledge could become over time one of the mainstays of research in psychology along with, among others, methodological rigor, honesty, and transparency. The usefulness of these two questions is explored by asking them with respect to one of the fields of research that among many others has come under scrutiny in the ensuing reproducibility debate (e.g., Pashler & Wagenmakers, 2012): *Research on social priming* (e.g., Bargh, Chen, & Burrows, 1996; Bargh & Chartrand, 1999).

## The Present ‘Crisis of Confidence’ in Psychology

### The problem is epistemology, not statistics

Much has been written and said about the “crisis of confidence” (Pashler & Wagenmakers, 2012) in psychology since the ominous years 2011 and 2012 when Diederik Stapel’s academic fraud was discovered, when Daryl Bem was able to publish a paper supposedly ‘proving’ human’s precognitive abilities in the JPSP (Bem, 2011; see also Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011), and when doubts were cast on the reproducibility of social priming effects (Doyen, Klein, Pichon, & Cleeremans, 2012). Consequently, initiatives such as the *Open Science Collaboration* (Open Science Collaboration, 2015) and the *Many Labs Project* (Klein et al., 2014) set out to further investigate the degree to which psychological studies can be replicated.

A closely related problem is the apparently frequent use of *questionable research practices* (QRPs; Fanelli, 2009; Leslie, Loewenstein, & Prelec, 2012; Martinson, Anderson, De Vries, 2005) such as *p-hacking* (making a statistically non-significant result appear significant; Simmons, Nelson, & Simonsohn, 2011) or *HARKing* (making up hypothesis after the results are known while pretending that they had been formulated in advance; e.g. Kerr, 1998) in the field of psychology. Taken together, these practices may make the current practice of significance testing in psychology more or less obsolete: With enough perseverance, *p-hacking*, and *HARKing*, researchers can create empirical evidence in favor of more or less any theoretical assumption (Simmons, Nelson, & Simonsohn, 2011; Gelman & Loken, 2013). As a consequence, a reader of scientific publications in the field of psychology apparently cannot assume with reasonable certainty that the reported research findings are ‘true’ in the sense that they could be replicated by independent researchers.

A number of authors have proposed statistical and methodological solutions to these problems such as using Bayesian instead of frequentist statistics (e.g., Marsman et al., 2017; Wagenmakers et al., 2017) or using stricter thresholds in null-hypothesis tests (Benjamin et al., 2017). However, several other authors (e.g., Holtz & Monnerjahn, 2017, Strack, 2017) reiterated Paul Meehl’s statement from 1997 that “The problem is epistemology, not statistics” (p. 394). Meehl’s point of departure here and in other

publications was still, of course, a statistical one: The way null-hypothesis tests are used in psychology does not put theories in “grave danger of refutation” (1978, p. 806). In contrast to the *hard sciences* such as physics, the theories or conjectures (if we want to use the term theory only for substantial and elaborated systems of knowledge) in *soft sciences* such as psychology most often do not yield point predictions in the sense that they predict a certain measurable numerical outcome (see also Meehl, 1967). In psychology, conjectures usually posit only that a given factor has some measurable influence on outcome variables. As a consequence—the aforementioned questionable research practices notwithstanding—even a random conjecture has in principle a 50% chance of being ‘verified’ in an empirical study given infinite sample sizes (unlimited statistical power): With increasing numbers of participants, the null-hypothesis is more and more likely to be rejected; when the number of participants converges towards infinity, the question is just whether the (almost certainly significant) effect goes in the right direction or not.

These and other considerations prompted Meehl to the conclusion that the established use of statistics in psychology leads to an overly optimistic appraisal of the truth status of psychological conjectures, while at the same time not enough attention is paid to potential alternative explanations. As a consequence, the growth of scientific knowledge in the field of psychology is not as fast or steady as it could be: By (more or less) only confirming what everyone believes to be true anyway, psychologists miss out on many opportunities to improve their theories as a consequence of research findings contradicting their assumptions. Similar arguments have been brought forward by a number of other authors before (e.g., Pettigrew, 1991; Meehl, 1990, 1997) and after (e.g., Earp & Trafimow, 2015; Holtz & Monnerjahn, 2017) the emergence of the current crisis in psychology.

One of the reasons for the apparent unwanted optimism with regard to empirical confirmation of one’s theory may be related to the strong publication pressure that emerged over the course of the “academic capitalism” which developed in the 20th century (Münch, 2014): Scientists increasingly have to *publish or perish* and to *get visible or vanish* (Doyle & Cuthill, 2015; Holtz, Deutschmann, & Dobewall,

2017) as means of surviving in the academic job market and of getting tenured positions. More than 90% of the publications in the field of psychology present evidence in favor of a theory (Fanelli, 2011; Sterling, Rosenbaum, & Weinkam, 1995; Yong, 2012). Alternative hypotheses and competing theories are only considered in a small number of cases (e.g., 21.6% and 11.4% respectively of 236 JPSP articles which were published between 1982 and 2005, according to Uchino, Thoman, and Byerly, 2010). Scientists more often than not build their careers around a certain ‘pet theory’ and their writings are mostly read and in turn cited by other scientists from the same research community. Such communities are built upon the assumption that a given theory is ‘true’ or at least useful in explaining relevant phenomena. Hence, being overly optimistic with regard to one’s pet theory can be a means of ensuring funding, publications, and hence survival in a highly competitive academic world (for an extensive analysis see Billig, 2013).

This attitude of using empirical research as a sales angle for a theory to be tendered in the academic market of ideas (and often enough in non-academic markets as well, such as consulting) was recently expressed in a very straightforward way by Daryl Bem in an interview with Slate magazine (Engber, 2017): “If you looked at all my past experiments, they were always rhetorical devices. I gathered data to show how my point would be made. I used data as a point of persuasion, and I never really worried about, ‘Will this replicate or will this not?’”. Whereas I believe that demanding ‘impartiality’ and ‘objectivity’ from researchers is illusory (see e.g., Holtz & Monnerjahn, 2017; Holtz & Odağ, 2018), I think nevertheless that the drawbacks of such a salesman-like approach to science are fairly obvious. Such an attitude is particularly detrimental when there are no serious alternative theories regarding the phenomena to be explained. Without competing theories, there is no competition on the market of ideas and no ‘checks and balances’ by scientists holding different views, which, according to philosopher Karl Popper (e.g., 1972) for example, are urgently needed to keep researchers’ optimism with regard to their theories at bay. It should be noted that Bem’s habit of using data to *show* his point—which he likely employed during most of his prodigious 50 plus-year career in social psychology—was only regarded as

problematic after he attacked a widespread common sense assumption: Human beings don’t have precognitive abilities. But how can we change such salesman-like attitudes?

### Changing the hearts and minds of researchers

The focus of the paper presented here will be on questions about if and how unwanted optimism in looking for confirmations of theories and finding these theories too easily corroborated in empirical studies can be addressed in and of itself. I am not going to discuss the specific statistical solutions Meehl and others have proposed for these problems. Without denying the importance of the related statistical debates, I believe that a change for the better is not only needed in terms of methodology, but also in terms of the general mindset that guides psychologists in conducting their research.

Ioannidis (2005) wrote in his seminal article on ‘false positives’ in science: “Diminishing bias through enhanced research standards and curtailing of prejudices may also help. However, this may require a change in scientific mentality that might be difficult to achieve” (p. 0701). Apart from statistical recommendations such as larger sample sizes and power analysis, Ioannidis also proposed to have researchers pre-register their intended studies whenever possible. This suggestion has been echoed by a number of scholars from the field of psychology (e.g., Lindsay, Simons, & Lilienfeld, 2016; van’t Veer & Giner-Sonolla, 2016), and several journals have implemented some form of a pre-registration policy since then (see e.g., Chambers, Feredoes, Muthukumarasway, & Etchells, 2014; Jonas & Cesario, 2016).

Apart from pre-registration, in their *manifesto for reproducible science*, Munafo and colleagues (2017) propose the ‘blinding’ of researchers as a means of protecting against cognitive biases (p. 2), for example in the form of not telling those who analyze the data which data points represent the experimental condition. They suggest further measures, such as educating researchers about effects of questionable research practices and defining guidelines for making data collection and analysis processes more transparent. Researchers should also be given incentive by funding agencies and publishers for following these open science recommendations.

I will argue in the next paragraphs in favor of making use of two simple questions that physicist John Platt formulated in an article in 1964 as a means of ensuring and accelerating progress in science in general. These two questions sum up in an easily understandable and accessible way the central ‘mantra’ of competing approaches in current epistemology: Scientific progress can only be defined as an advancement beyond existing knowledge, and we can learn most of all from our mistakes. I will call such an approach to the growth of (scientific) knowledge *critical thinking* or a “critical approach” (Popper, 1962, p.51). The advantage of my proposal is that it can be implemented easily with any kind of study. Many of the aforementioned methodological and statistical measures more or less follow directly from paying more attention to competing theories and the falsification of results. Hence, I propose first to tackle the ‘hearts and minds’ of researchers by introducing a simple behavior pattern which will prepare the ground for the methodological suggestions mentioned above for improving the reproducibility of psychological science. I believe that if only a small group of researchers would start asking these two questions on a regular basis, there is a good chance that critical thinking could become just as commonplace as preregistration and replicability research have become over the last years.

### Platt’s ‘Two Questions’

Platt’s (1964) main point of departure was that, from his perspective, progress (at least during the 1960s) apparently happened faster and more steadily in certain branches of science such as high-energy physics and molecular biology than in others. He attributed these differences to the frequent use of *strong inference* in these disciplines. According to him, strong inference entails the following steps:

- 1) Devising alternative hypotheses;
- 2) Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
- 3) Carrying out the experiment so as to get a clean results;
- 1) Recycling the procedure, making subhypotheses or sequential hypotheses to refine the

possibilities that remain and so on. (p. 347)

Of course, this idea is not new. Actually, the concept of “strong inference” by means of comparing competing theories had already been introduced by the geologist Chamberlin in 1890, and the idea of an *experimentum crucis* as a means of testing competing theories per se extends at least back to Bacon’s “*Novum Organon Scientiarum*” in the 17th century (Bacon, 1620). It must also be noted that Platt’s accounts of the history of strong inference and its use in modern science has been criticized for historical inaccuracies and for over-simplifying the issues that are related to devising crucial experiments (e.g., O’Donohue & Buchanan, 2001). So why should we return to his article, which was written over 50 years ago by a perhaps overly enthusiastic physicist, as a potential partial remedy for the current crisis in psychology?

Platt’s analysis of the problems in many scientific fields corresponds to my (of course limited) experience in the field of psychology: Scientists—when they are asked what science ideally should be like—usually know that they are supposed to be critical, employ rigorous tests of their theories, and compare theories whenever possible. However, in our daily lives as scientists, our minds are occupied with other things: For example, we have to rapidly publish our findings—lots of them—to keep our careers going, and we have lots of other everyday duties to fulfill that may prevent us from employing the scientific rigor that we know is needed. As Platt puts it:

How many of us write down our alternatives and crucial experiments every day, focusing on the *exclusion* of a hypothesis? We may write our scientific papers so that it looks as if we had steps 1, 2, and 3 in mind all along. But in between, we do busywork. We become “method-oriented” rather than “problem-oriented.” We say we prefer to “feel our way” toward generalizations. We fail to teach our students how to sharpen up their inductive inferences. (p. 348; emphasis as in the original)

However, the strongest part of Platt’s paper is in my opinion his description of “aids” (p. 352) for the implementation of strong inference in the daily scientific practice. How can we—in the middle of our everyday struggles—enforce a critical mindset upon ourselves, our students, and our colleagues? Platt proposes two simple questions that can and should

be asked about any scientific studies as means of employing a “yardstick” (p. 352) for the study’s effectiveness:

[Question 1:] But sir, what experiment could disprove your hypothesis?”; or, on hearing a scientific experiment described,

[Question 2:] But sir, what hypothesis does your experiment disprove?

This goes straight to the heart of the matter. It forces everyone to refocus on the central question of whether there is or is not a testable scientific step forward. (p. 352)

He continues:

If such a question were asked aloud, many a supposedly great scientist would sputter and turn livid and would want to throw the questioner out, as a hostile witness! Such a man is less than he appears, for he is obviously not accustomed to think in terms of alternative hypotheses and crucial experiments for himself; and one might also wonder about the state of science in the field he is in. But who knows?—the question might educate him, and his field too! (p. 352)

In the next section, I will discuss the epistemological importance of these two questions as well as criticism of the empiricist ‘success formula’ that Platt had presented in his paper.

### **The Epistemological Importance of Platt’s “Two Questions”**

First, it should be noted that to Platt, these two questions are in fact just one question. Following the empiricist tradition, Platt believed in the *experimentum crucis* as the driving factor behind scientific progress. In this sense, asking what kind of evidence could go against one’s theory is always the same as asking for evidence that would support another theory, because the effectiveness of an *experimentum crucis* depends on the possibility of identifying relevant competing theories as well as deciding which one is ‘wrong’ and which one is ‘right’. Probably like every ‘Baconian’ ever since the 17th century, Platt may be oversimplifying matters here: First of all, it is not always possible to identify all of the relevant competing theories. Actually, in some areas of science, we may be faced with the fact that we have entered uncharted territory insofar as there is at that point no theory yet that could explain the phenomena in question. Furthermore, the idea of

setting up studies that once and for all could clarify which theory is right and which is wrong is probably naive in ignoring underdetermination, or the ‘Quine-Duhem Thesis’ (Harding, 1976): Research findings are not only influenced by the theoretical mechanisms that we want to test, but also by innumerable so-called *auxiliary hypotheses*, ranging from rather trivial assumptions such as ‘our measurement instrument worked’ to serious hitherto unknown alternative explanations. For example, in 1906, physicist Pierre Duhem mathematically proved for a subfield of physics (Newton’s law of universal mutual gravitation) that the number of such auxiliary hypotheses is necessarily infinite. Subsequently, philosopher W. V. O. Quine (1951) generalized Duhem’s argument to more or less everything that can be known to human beings. The *experimentum crucis* idea also becomes critical whenever a theory does better at explaining certain phenomena in certain areas while a competing one offers better explanations in other areas.

However, Platt’s two questions also make lots of sense for those who are aware of the limits of empiricism. For example, critical rationalists in the tradition of Karl Popper assume that no empirical method whatsoever can clarify theoretical questions once and for all. All that empirical methods can do is to discover inconsistencies between theoretical predictions and empirical results (*falsification*). They can thus point towards ways in which theories can be improved. Over time, employing such a critical mindset towards theories and working on ways of improving them lead to a growth of knowledge in an evolutionary sense (Popper, 1972): Theories are continuously replaced with better theories, and this process refines our understanding of the world, although we will never know with certainty whether or not our theories are true in a metaphysical sense.

Consequently, even without considering a competing theory, just asking for empirical results that would go against a researcher’s expectations can in itself be an important part of the research process: Scientific progress is possible to the critical rationalist only through discovery of such inconsistencies. For the second step, the replacement of theories with ‘better’ theories, comparing and critically evaluating theories is pivotal for obvious reasons. Hence, Platt’s ‘question’ can be put to use in critical rationalism only after making ‘two questions’ of the one.

In the next paragraph, I will formulate a more generalized version of these questions that is particularly suited for use in all branches of psychology. My objective here is again to facilitate the asking of these questions on a regular basis as a means of making critical thinking or a critical approach “mainstream” in the field of psychology. I believe that without the ballast of a formalized empiricism, Platt’s two questions can be even stronger tools to foster critical thinking in the field of psychology.

### The two questions revisited

The way Platt formulated his questions may scare off some psychologists through the use of strong terms such as ‘disprove’ and by explicitly mentioning ‘experiments’ as the method of choice. Because I believe that the same epistemological principles guide qualitative as well as quantitative research traditions in psychology (see Holtz & Odağ, 2018), I would like to introduce a more generalized wording of Platt’s two questions that in my opinion is better suited for use in a multifaceted field such as psychology.

New wording for the first question should address the identification of possible competing theories, for example in the following form: “Are there any reasons to expect a different outcome than the one you expect?” If then reasons are provided to expect a different outcome, the follow-up sub-question should be: “To what extent can our study provide arguments in favor or against the competing assumptions?” If this question is answered in the negative, one should think about ways to refine the study.

The second question is meant to make the researcher aware that science is not only about finding confirmations of one’s beliefs, but rather about critically testing them. The researcher must remain aware that more often than not, inconsistencies between predictions and expectations drive scientific progress: “Which outcomes would clearly contradict your assumptions?” or another way to word it would be, “Which outcomes could cast doubt on confidence in your underlying assumptions?”. This question first of all has the purpose of making the researcher aware of possible discrepancies between predictions and observations. This results in the critical mindset that in critical rationalism is of the two driving forces behind scientific progress (the other one is “intuition or imagination”; Popper 1979,

p. 167). It may be followed up by the question “And what consequences is it going to have if you get results that go against your expectations?”.

In the following paragraphs I will attempt to demonstrate the usefulness of these two questions with their follow-up questions for the field of psychology by asking them not about a single study, but about what could maybe be called a *research programme* (Lakatos, 1978) in social psychology: Research on *Social Priming* (e.g., Bargh, Chen, & Burrows, 1996; Bargh & Chartrand, 1999). I think a very similar argument could be made for other research programmes such as *ego depletion research* (in the tradition of Baumeister, Bratslavsky, Muraven, & Tice, 1998; for a critical perspective see e.g. Carter & McCullough, 2014) or *power posing research* (in the tradition of Carney, Cuddy, & Yap, 2010; for a critical perspective see e.g. Ranehill et al., 2015).

It is important to keep in mind that I do not want to propagate the idea of strong inference in the sense of Platt (1964). I just want to use his two questions to explore in the form of a thought experiment how psychology would benefit from more critical thinking or a critical approach.

## The Case of Social Priming

### The rise and fall of social priming research

In psychology, the word *priming* usually designates effects of exposure to a stimulus (such as a word, a bodily sensation, or an observation) on a subject’s responses in a situation subsequent to the stimulus exposure. Early research on priming focused primarily on how reading certain words had an effect on the perception and processing of subsequent associated and/or semantically related words (e.g., Meyer & Schvaneveldt, 1971; Meyer & Schvaneveldt, 1976; Neely, 1977). The term *social priming* is most often used as an umbrella term for different kinds of priming in the form of *unconscious activation of social categories* (such as old, polite, rude, ...), resulting in behavioral tendencies that are based on those respective schemes, role concepts, or stereotypes (e.g., Molden, 2014). The term social priming has been used by, among others, Daniel Kahneman in his concerned open letter (2012) to the “students of social priming”. Kahneman had previously devoted a part of his best-selling book *Thinking, Fast and Slow* (2011) to social priming research.

In a seminal study that paved the way for a large number of other publications with variations on the putative underlying phenomenon of social priming, Bargh, Chen, and Burrows (1996) found, among other results, behaviors that allegedly stemmed from social priming: Participants in laboratory experiments were more likely to interrupt a conversation between an experimenter and a confederate when, prior to the conversation, they had solved verbal problems (so called ‘scrambled-sentence tasks’) which included words that were related to rudeness (e.g., bold, aggressively, rude, ...) than when the problems included words that were related to politeness or were ‘neutral’ in this regard. In another experiment, participants walked down a hallway significantly more slowly when they had previously solved problems which included words that were related to old age (e.g., old, Florida, grey, ...) than when the problems included neutral words or words that were associated with youth.

Although the paper by Bargh and colleagues (1996) has been quoted literally thousands of times, according to Doyen, Klein, Pichon, & Cleeremans (2012), only two partially successful replication studies were published between the years 1996 and 2012 (Aarts & Dijksterhuis, 2002; Cesario, Plaks, & Higgins, 2006). In a first experiment of their own, Doyen and colleagues were unable to replicate the *primed elder-walking experiment* of Bargh and colleagues (1996); in a second study, they succeeded at producing similar effects as those in the original study—but only if the experimenter was aware of the hypothesis and not if s/he was ‘blind’ towards the expected outcome. After initial discussions regarding differences between the setups of the experiments and after what some commentators perceived to be attacks by Bargh and colleagues against the quality and credibility of Doyen and colleagues’ paper (Yong, 2012), the controversy regarding the replicability of social priming effects continues until today (e.g., Weingarten, Chen, McAdams, Yi, Hepler, & Albarracín, 2016; Schimmack, Heene, & Kesavan, 2017; see also Daniel Kahneman’s related comment as reported in McCook, 2018).

In the following paragraphs, I will ask Platt’s two questions with a view to the social priming research programme (Lakatos, 1978) as a whole, and I will discuss potential consequences: [Question 1] Are there

reasons to expect a different outcome, and [Question 2] what could some consequences of unexpected findings be?

### **Are there reasons to expect a different outcome?**

In their 1996 paper, Bargh and colleagues argue (p. 230–231) that whereas it is widely accepted that attitudes, emotions, and self-concepts can be affected in an unconscious way by means of activating schemes and scripts and the like, “behavioral responses to the social environment are [usually believed to be] under conscious control” (p. 230). They continue by quoting two authors (Fiske, 1989; Devine, 1989) who acknowledged that behavior may be affected by automatic processes, for example, in the form of the unconscious activation of stereotypes; however, Fiske and Devine both made the claim that human beings can still consciously decide to overcome such behavioral tendencies and decide not to act in accordance with their prejudice. Social priming research makes the contrary claim that behavioral responses to automatic cognitive processes are neither mediated by attitudes and emotions nor can they be *overruled* through higher cognitive instances because they happen unconsciously.

But who would doubt that such cases of unconscious behavior affected by priming exist? Bargh and colleagues seem to assume according to the aforementioned quote that some imagined opponent would maintain that there are none, and that there cannot be any cases in which behavior is affected in an unconscious and automatized way through scheme activation and the like. In formal logic, such a statement could have the form of the implication  $A \Rightarrow B$  (read: ‘if A, then B’ or ‘whenever A is the case, B is going to be the case as well’) with A being a behavioral response and B being some degree of conscious control. In this case, in accordance with the so-called *modus tollens*, a singular observation of A and non-B would make the  $A \Rightarrow B$  clause false. But are there really opponents who completely rule out the possibility that there can be a behavioral response without conscious control?

At least Fiske (1989) and Devine (1989), the two authors whom Bargh and colleagues (1996) quote as proponents of a conscious cognitive behavioral control mechanism (p. 231), would probably not subscribe to any statement positing the impossibility of unconscious effects on behavior in a radical form. To

me, it rather seems that unconscious mechanisms have been one of the defining features of psychology as a discipline from the days of the psychoanalysts and the gestalt theorists to modern social psychology.

But let us imagine such a radical opponent, for example, in the form of a stubborn economic rationalist or a theologian who steadfastly believes in human beings' *free will*, and who refuses to acknowledge the possibility that some of our behavior is the consequence of automatized and unconscious psychic processes: Is there any chance that such a person could be convinced through studies such as the 'primed elder-walking experiment'? Probably not. One reason is that the experiments by Bargh and colleagues were statistically 'underpowered' (Schimmack, Heene, & Kesavan, 2017; Weingarten, Chen, McAdams, Yi, Hepler, & Albarracín, 2016). This means that given the average effect size, the sample sizes in the experiments (ranging from 30 to 34) were not sufficient to ensure significant results with reasonable certainty (e.g., >80%). A mean-spirited opponent would immediately pick on this deficiency. Furthermore, the description of the procedure follows the conventions of the day, but a hostile adversary would maybe point out that the information is not sufficient to enable exact replication of the study in question (see e.g. Stark, 2018). A steadfast opponent would also point out that the study was not pre-registered, and that neither the data nor analysis procedures were made public. Furthermore, Bargh and his colleagues would have to state how many other experiments conducted by them failed to yield reliable results in a predicted direction and so were relegated to the 'file drawer', rather than being published explicitly along their 'successful' experiments. An opponent could also point to the artificiality of such psychological laboratory experiments and demand evidence from studies using more 'natural' behavior data (McGuire, 1973 & 2004). All in all, to convince such an opponent, the design as well as the reporting standards would probably have to be much stricter.

In a later publication, Bargh and Chartrand (1999) presented a stronger version of their thesis based on the studies in the 1996 paper and a few related studies:

Our thesis here—that most of a person's everyday life is determined not by their conscious inten-

tions and deliberate choices but by mental processes that are put into motion by features of the environment and that operate outside of conscious awareness and guidance—is a difficult one for people to accept. (p. 462)

This is of course a more provocative thesis that will most definitely meet opposition among at least a few psychologists and other social scientists arguing in favor of a more humanistic image of a human being as an at least partly rational being that is able to defy its more animalistic tendencies. The statement in its strong form also bears relevance for ethical and jurisdictional debates: Is a human being really responsible for her actions if 'most' of her life is 'determined' by environmental forces operating 'outside of conscious awareness and guidance'?

But do Bargh and colleagues' findings here and elsewhere support this far-reaching conclusion? No. The fact that it is possible to create an experiment in which unconscious factors affect behavior does not allow for generalized conclusions about the importance of these processes in everyday life or to the percentage of everyday decisions that are affected or even determined (sic) by uncontrollable unconscious forces. This is all of course assuming that social priming experiments are in fact replicable.

Thus, it seems that Bargh and colleagues are first creating a strawman by making the untenable claim that behavior is *always* under conscious control. Then they use their findings to propagate a much more far-reaching theory of the *conditio humana* as a miserable being that, evoking Freud, is not even master in its own house.

Of course, research questions regarding the degree to which behavior is under conscious control are valuable. But in this case, just creating an experiment that does (or does not) show that unconscious effects are possible is not enough. Different kinds of studies comparing behavioral reactions systematically in different scenarios (in the sense of the aforementioned strong inference) would be needed to allow for this kind of generalization (see also McGuire, 1973 & 2004). Such studies would also have to be adequately powered in terms of the number of subjects to allow for these kinds of conclusions.

### **What if you don't find what you expected?**

Is there a possible research outcome that would eventually cause Bargh and colleagues to abandon their idea of unconscious automatized effects on behavior? Does the strawman mentioned earlier have any theoretical chance at all to defeat his opponents? I don't think so. One reason is that it is all but impossible statistically to clearly demonstrate the absence of any effect (e.g., Cohen, 1994).

Of course, the case is different for the stronger statement that such unconscious effects determine most of our daily lives. But here as well, the question of whether or not it is possible to demonstrate unconscious effects in an experiment seems to be by and large irrelevant. Obviously, there are a large number of studies showing the rather trivial fact that conscious processes can 'override' automatic behavior tendencies (e.g., in Devine, 1989, mentioned beforehand). But for estimating the relevance of any findings for everyday processes, a more comprehensive approach would again be needed.

So what is the theoretical consequence of Bargh and colleagues' studies being replicable or not? Would any imaginable outcome of a large-scale replication study have an effect on our understanding of the world and human nature? Maybe a failure to replicate Bargh and colleagues' findings could be used as an argument that, after all, it is not so easy to manipulate human beings' minds. But actually the relevance of this argument could be more related to the large degree of publicity that Bargh and colleagues' studies received than to the empirical evidence per se. In itself, the fact that an experiment 'does not work' (does not yield the expected results) does of course not rule out the fact that another experiment could be designed that would 'work', finally demonstrating the intended effect.

### **Newell's (1973) infinite yes-or-no game and the question why social priming research has become so popular**

Applying Platt's two questions to classical social priming studies seems to indicate that Bargh and colleagues' studies present little relevant empirical evidence with regard to the most relevant question as to what extent human beings can consciously control their behavior. As I have explained, no psychologist would seriously rule out the possibility

that there may be unconscious effects of automatized cognition on behavior, and the experimental setup of the studies does not really allow for any generalization in the sense of estimating the degree to which everyday behavior is affected by conscious and unconscious processes. The small sample sizes, the relatively loose research protocol (no preregistration), and the relatively sparse information about the procedures—which were however absolutely in line with the conventions of the day—are factors that also limit the study's capacity to convince a stubborn opponent of potential priming effects of their existence. In order to provide empirical evidence supporting the stronger thesis that 'most' of our behavior is 'determined' by unconscious processes, a wider range of studies systematically comparing conscious and unconscious influences and aiming at discovering the limits of the respective underlying assumptions would be more helpful than just repeating experiments that show that some unconscious automatized effect can be elicited. But why have variations of exactly these studies then become so popular?

One reason may be related to the naive positivistic idea that obtaining the expected result in a thoroughly controlled scientific experiment 'proves' the existence of the phenomenon in question. This claim has been by and large discredited in philosophy, for example, because of the aforementioned issues with underdetermination (Quine, 1951). Even given a reproducible stable effect, someone could come along at any time with a better explanation of the observed phenomena in question and show hitherto unknown evidence falsifying the original assumptions (e.g., Popper, 1962).

Thinking about social priming's popularity, I am also reminded of Newell's (1973) brilliant analysis of the issues in psychological research in the 1970s. He argued that psychologists too often study complex questions by means of reducing them to a *yes or no* type question. Examples would be the nature vs. nurture debate or the debate over conscious vs. unconscious information processing already ongoing at that time (Newell lists 24 such yes or no questions on p. 288). The true answer to all these questions is most likely one starting with "it all depends", and answering them in a productive way would actually require complex theories, elaborated research designs, and strong inference (as outlined in Platt, 1964). However, asking complex questions in a yes or

no fashion and only counting confirmations of the respective theories sets the stage for an infinite game: Apparent opponents produce an endless chain of evidence in favor of their own stance regarding such a yes or no question without ‘hurting’ each other and without necessarily producing anything resembling a growth of knowledge. The right question to ask here in order to achieve progress would be “to what extent is your research potentially able to convince an opponent of your point of view?” Hopefully productive discussions would then be facilitated among proponents of different views.

### Conclusion

Science is not about finding your assumptions confirmed and finding ways to sell your ideas to an audience. Instead, it is (or should be) about critically examining your beliefs and correcting them whenever they do not correspond with empirical observations, and about being willing to give them up whenever there is a better explanation for the phenomena in question. This is what I call *critical thinking* or a *critical approach* to the growth of scientific knowledge. I assume that most psychologists would agree with these statements. Still, the question is to what extent scientists also (can) act according to these principles in a globalized capitalist academic market where productivity and public attention determine a scientist’s career. It should be the task of scientific organizations such as scientific associations editing prestigious journals to help scientists to act in accordance with what they know would be the right thing to do. Reward structures should be established that reward good scientific practices, such as taking into account alternative explanations, exploring the limits of one’s assumptions, and being open to report also unexpected results (e.g., Munafò et al., 2017). Just as large-scale replication projects, preregistration, and open science have been popularized by small groups of researchers continuing to make their point, I hope that critical thinking could become a mainstay of psychology as well, if enough researchers begin to ask Platt’s two questions on a regular basis.

In view of the most recent crisis in psychology, I think it is also fair to demand from those who are active in teaching scientific methods to students to

put as much emphasis on developing critical thinking as they put on teaching thorough methodological knowledge. For those of us who work on scientific textbooks and who convey scientific knowledge to the wider public, I think that the question should be asked whether a salesman-like attitude of (over-) selling the benefits of scientific research is really a way to sustain worthwhile scientific activity.

Platt’s Two Questions indeed can and should be asked with regard to any scientific study in the field of psychology, and I hope that I have provided some arguments for my cause. At least I have the impression that the scientific output throughout my scientific career might have been more relevant and interesting if I had asked these questions about my own work on a regular basis. Of course, asking the questions does not entail having or prescribing an answer. I am aware that my discussion of social priming research is to some degree provocative and will evoke criticism from those who are more knowledgeable in this area of psychology than I am. Maybe they can convince me that these questions can or should be answered differently for social priming research.

The most critical part of this paper is probably not the question whether Platt’s two questions should be asked on a regular basis—I suppose there will not be much opposition to this thesis. Instead, it is my claim that these questions have not been asked often enough in psychology so far. Still, I hope that I could at least make the point that the methodological recommendations that are meant to counter the current crisis (e.g., Benjamin et al., 2017; Munafò et al., 2017) follow immediately from a more critical approach to scientific research.

Eventually, critical thinking could become just as ‘mainstream’ as preregistration and large-scale replication studies have become over the last several years. And in the end ‘market forces’ could themselves contribute to a critical culture in psychology once a critical degree of popularity has been reached: As soon as scientists have to demonstrate their ability to think critically in order to, for example, obtain a tenured position, a significant incentive will have been created. Consequently, scientists may then have to document, for example, how many hypotheses they have falsified or the number of occasions on which an empirical finding had made them revise the theory to be tested. Falsifications of null-hypotheses would only count here whenever they

constitute a critical test of the theory to be tested in the sense that other researchers would be willing to defend the null-hypothesis, as is the case, for example, in para-psychology. Here, most scientists hold the view that psy-phenomena do not exist and falsifying this hypothesis (of course, in a reproducible and replicable way) would be a substantial advancement of scientific knowledge. In the previously discussed example of social priming, the case is different and—as I argued in the previous paragraphs—a rejection of the null-hypothesis may not mean much from an epistemological point of view.

I would consider it an ideal outcome of this paper if there were a kind of movement in the direction of simply asking these two questions on a regular basis and studying the reactions. I would expect some respondents to indeed turn 'livid', whereas I hope that others would start heading towards more critical thinking. Social media could perhaps facilitate the exchange of experiences with reactions to these questions.

#### Author Contact

Dr. Peter Holtz. Schleichstr. 6, 72074 Tübingen, Germany; [p.holtz@iwm-tuebingen.de](mailto:p.holtz@iwm-tuebingen.de). Orcid: <https://orcid.org/0000-0001-7539-6992>.

#### Conflict of Interest and Funding

I have no conflict of interest to declare. My research is funded since 2018 by the Leibniz Association, Germany (Leibniz Competition 2018, funding line "Collaborative Excellence", project SALIENT [K68/2017]).

#### Author Contributions

I am the sole author of this article.

#### Acknowledgements

Several in-text comments by David E. Meyer on the first version of this article were integrated into this revised version of the original manuscript. I would like to thank Professor Meyer for his invaluable contribution to the paper!

I also want to thank reviewer Michael Olson and the editors Åse Innes-Ker and Rickard Carlsson for

their valuable feedback and their help during the review process!

#### Open Science Practices

This is a conceptual article without data, materials or analysis that could have been preregistered. The entire editorial process, including the open reviews, are published in the online supplement.

#### References

- Aarts, H., & Dijksterhuis, A. (2002). Category activation effects in judgment and behaviour: The moderating role of perceived comparability. *British Journal of Social Psychology*, 41(1), 123-138.
- Bacon, F. (1620). *New Organon*. Available from: [http://www.constitution.org/bacon/nov\\_org.htm](http://www.constitution.org/bacon/nov_org.htm); last retrieved June 2018.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71(2), 230.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American psychologist*, 54(7), 462.
- Baumeister, R. F.; Bratslavsky, E.; Muraven, M.; Tice, D. M. (1998). Ego depletion: Is the active self a limited resource?. *Journal of personality and social psychology*, 74(5), 1252-1265.
- 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.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Cesarini, D. (2018). Redefine statistical significance. *Nature human behaviour*, 2(1), 6.
- Billig, M. (2013). *Learn to write badly: How to succeed in the social sciences*. Cambridge: Cambridge University Press.
- Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological science*, 21(10), 1363-1368.
- 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, 823.
- Cesario, J., Plaks, J. E., & Higgins, E. T. (2006). Automatic social behavior as motivated preparation to interact. *Journal of personality and social psychology*, 90(6), 893.
- Chambers, C. D., Ferdeous, 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.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American psychologist*, 49(12), 997.
- Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, 15(366), 92-96.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56, 5-18.
- Dewey, J. (1903/2004). *Democracy and education*. Mineola, NY: Dover.
- 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.
- Duhem, P. (1954/1906). *The Aim and structure of physical theory* (Transl. P. P. Wiener). Princeton, NJ: Princeton University Press.
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in psychology*, 6, 621.
- Engber, D. (2017). Daryl Bem proved ESP is real: Which means science is broken. *Slate*. Online document available from <https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html>; last retrieved May 2020.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS one*, 4(5), e5738.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.
- Feynman, R. P. (1974). Cargo cult science. *Engineering and Science*, 37(7), 10-13.
- Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.): *Unintended thought*. New York: Guilford Press, 253-283.
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University. Online document available at: [http://www.stat.columbia.edu/~gelman/research/unpublished/p\\_hacking.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf); last retrieved June 2018.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, 351(6277), 1037-1037.
- Harding, S. (Ed.). (1976). Can theories be refuted?: *Essays on the Duhem-Quine thesis* (Vol. 81). New York: Springer.
- Holtz, P., Deutschmann, E., & Dobewall, H. (2017). Cross-cultural psychology and the rise of academic capitalism: Linguistic changes in CCR and JCCP articles, 1970-2014. *Journal of cross-cultural psychology*, 48(9), 1410-1431.
- Holtz, P., & Monnerjahn, P. (2017). Falsificationism is not just 'potential' falsifiability, but requires 'actual' falsification: Social psychology, critical rationalism, and progress in science. *Journal for the theory of social behaviour*, 47, 348-362.
- Holtz, P., & Odağ, Ö. (2018). Popper was not a positivist: Why critical rationalism could be an epistemology for qualitative as well as quantitative social scientific research. *Qualitative research in psychology*, Advance online publication. Online document available from: <https://doi.org/10.1080/14780887.2018.1447622>; last retrieved May 2020.
- 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.
- Jonas, K. J., & Cesario, J. (2016). How can preregistration contribute to research in our field?. *Comprehensive Results in Social Psychology*, 1, 1-7.
- Kahneman, D. (2012). *A proposal to deal with questions about priming effects*. Online document available from: [https://www.nature.com/polopoly\\_fs/7.6716.1349271308!/sup](https://www.nature.com/polopoly_fs/7.6716.1349271308!/sup)

- [pinfile/Kahneman%20Letter.pdf](#); last retrieved June 2018.
- Klein, R.A., Ratliff, K.A., Vianello, M., Adams Jr, R.B., Bahnik, Š., Bernstein, M.J., Bocian, K., Brandt, M.J., Brooks, B., Brumbaugh, C.C., & Cemalcilar, Z. (2014). Investigating variation in replicability: A “many labs” replication project. *Social psychology*, 45(3), 142-152.
- Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge, UK: Cambridge University Press.
- Lindsay, D. S., Simons, D. J., & Lilienfeld, S. O. (2016). Research preregistration 101. *APS observer*, 29(10). Online document available from <https://www.psychologicalscience.org/observer/research-preregistration-101>; last retrieved June 2018.
- Marsman, M., Schönbrodt, F. D., Morey, R. D., Yao, Y., Gelman, A., & Wagenmakers, E. J. (2017). A Bayesian bird's eye view of ‘Replications of important results in social psychology’. *Royal Society open science*, 4(1), 160426.
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737.
- McCook, A. (2018). “I placed too much faith in underpowered studies:” Nobel Prize winner admits mistakes. Online document available from <https://retraction-watch.com/2017/02/20/placed-much-faith-underpowered-studies-nobel-prize-winner-admits-mistakes/>; last retrieved June 2018.
- McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of personality and social psychology*, 26(3), 446.
- McGuire, W. J. (2004). A perspectivist approach to theory construction. *Personality and social psychology review*, 8(2), 173-182.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of science*, 34, 103-115.
- 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.
- Meehl, P. E. (1990). Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it. *Psychological inquiry*, 1(2), 108-141.
- Meehl, P. E. (1997). The Problem Is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In: L. L. Harlow, S. A. Mulaik, & J. H. Steiger (eds): *What If There Were No Significance Tests?*. London: Routledge, 393-425.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90(2), 227.
- Meyer, D. E., & Schvaneveldt, R. W. (1976). Meaning, memory structure, and mental processes. *Science*, 192(4234), 27-33.
- Münch, R. (2014). *Academic capitalism: Universities in the global struggle for excellence*. London: Routledge.
- Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1, 0021.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general*, 106(3), 226.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (ed.): *Visual Information Processing*. New York: Academic Press, 283-308.
- Molden, D. C. (2014). Understanding priming effects in social psychology: What is “social priming” and how does it occur?. *Social cognition*, 32(Supplement), 1-11.
- O'Donohue, W., & Buchanan, J. A. (2001). The weaknesses of strong inference. *Behavior and philosophy*, 29, 1-20.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence?. *Perspectives on psychological science*, 7(6), 528-530.
- Pettigrew, T. F. (1991). Toward unity and bold theory: Popperian suggestions for two persistent

- problems of social psychology. In C. W. Stephan, W. G. Stephan, & T. Pettigrew (eds.): *The future of social psychology*. New York: Springer, 13-27.
- Platt, J. R. (1964). Strong inference. *Science*, 146(3642), 347-353.
- Popper, K. R. (1934/1959). *The logic of scientific discovery*. London: Routledge (German original 1934).
- Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge*. Basic Books: New York & London.
- Popper, K. (1972). *Objective knowledge: An revolutionary approach*. Oxford: Oxford University Press, 246-284.
- Popper, K. R. (1979). *Three worlds*. Ann Arbor, MI: University of Michigan Press.
- Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological science*, 26(5), 653-656.
- Quine, W. V. O. (1951). Two dogmas of empiricism. *The philosophical review*, 60(1), 20-43.
- Schimmack, U., Heene, M., & Kesavan, K. (2017). *Reconstruction of a train wreck: How priming research went off the rails. Replicability-Index: Improving the Replicability of Empirical Research*. Online document available from <https://replicationindex.wordpress.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-off-the-rails/>; last retrieved June 2018.
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual review of psychology*, 69, 487-510.
- 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.
- Stark, P. B. (2018). Before reproducibility must come preproducibility. *Nature*, 557(7707), 613.
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The american statistician*, 49(1), 108-112.
- Strack, F. (2017). From Data to Truth in Psychological Science. A Personal Perspective. *Frontiers in psychology*, 8, 702.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59-71.
- Uchino, B. N., Thoman, D., & Byerly, S. (2010). Inference patterns in theoretical social psychology: Looking back as we move forward. *Social and personality psychology compass*, 4(6), 417-427.
- van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of experimental social psychology*, 67, 2-12.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. (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.
- Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological bulletin*, 142(5), 472.
- Wagenmakers, E. J., Verhagen, A. J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. In S. O. Lilienfeld & I. D. Waldman (Eds.): *Psychological science under scrutiny: Recent challenges and proposed solutions*. New York: Wiley, 123-138.
- Yong, E. (2012). Bad copy. *Nature*, 485(7398), 298.