ORIGINAL RESEARCH

Social sciences in crisis: on the proposed elimination of the discussion section

Philipp Schoenegger[1] 📖 · Raimund Pils[2]

Abstract

The social sciences are facing numerous crises including those related to replication, theory, and applicability. We highlight that these crises imply epistemic malfunctions and affect science communication negatively. Several potential solutions have already been proposed, ranging from statistical improvements to changes in norms of scientific conduct. In this paper, we propose a structural solution: the elimination of the discussion section from social science research papers. We point out that discussion sections allow for an inappropriate narrativization of research that disguises actual results and enables the misstatement of true limitations. We go on to claim that removing this section and outsourcing it to other publications provides several epistemic advantages such as a division of academic labour, adversarial modes of progress, and a better alignment of the personal aims of scientists with the aims of science. After responding to several objections, we conclude that the potential benefits of moving away from the traditional model of academic papers outweigh the costs and have the potential to play a part in addressing the crises in the social sciences alongside other reforms. As such, we take our paper as proffering a further potential solution that should be applied complimentarily with other reform movements such as Open Science and hope that our paper can start a debate on this or similar proposals.

Keywords Philosophy of the social sciences · Science reform · Open science · Incentive structures · Adversarial collaboration

Philipp Schoenegger and Raimund Pils have contributed equally to this work.

✉ Philipp Schoenegger
   contact.schoenegger@gmail.com

   Raimund Pils
   raimund.pils@sbg.ac.at

1 University of St Andrews, St Andrews, UK

2 University of Salzburg, Salzburg, Austria

🖄 Springer

1 Introduction

Social sciences are currently facing several interrelated crises, such as the replication crisis, the theory crisis, the applicability crisis, the generalizability crisis, and the validity crisis. These crises potentially threaten numerous aspects of the social scientific programme and the public perception of social science (Benessia et al., 2016; Hendriks et al., 2016). To this date, numerous potential solutions have already been proposed, ranging from changes in scientific norms to statistical training. Our focus is to present an additional, as-of-yet unrecognized potential solution, namely directly reworking the structure of academic papers. Specifically, we propose eliminating the discussion section from research papers. Our central claim is that eliminating discussion sections might improve the social science primarily because authors will have less opportunities for presenting their research in a biased way, leading to adverse downstream effects. We argue that removing the discussion section from papers and instead outsourcing it to independent discussion papers might eliminate many of the cognitive biases that make the discussion section problematic while not presenting substantial new costs. We further claim that this proposal can draw on several additional upsides, such as benefiting from the division of labour and from an adversarial mode of scientific progress. We see this proposal as working alongside, and not replacing, other reform efforts, and hope that the present paper kickstarts a debate on the merits and costs of the current structure of academic articles.

This paper is structured as follows. In Sect. 2 we briefly introduce the crises facing the social sciences and discuss current solutions. In Sect. 3 we show why these crises matter for social science and society more generally and how this points towards two distinct challenges. In Sect. 4 we develop our proposed solution of eliminating the discussion section and defend it against objections.

2 Social science in crisis

In this section we give a short upshot of the so-called 'crises' facing the social sciences. We focus on three central ones: the replication crisis, the theory crisis, and the applicability crisis; though what we say also applies to several others, such as the theory crisis and the validity crisis. Furthermore, our paper does not aim to address fraud per se, we are focused on less severe but still essential issues that arise in academic conduct.

Issues related to the replication crisis have been investigated thoroughly by social scientists, metascientists, and philosophers (e.g., Anvari and Lakens, 2018; Fletcher, 2021; Flis, 2019; Lilienfeld and Strother, 2020; Wiggins & Christopherson 2019). This debate was kickstarted in response to a replication failure of social priming findings (e.g., Bargh et al., 1996), which triggered several large-scale collaborative replication attempts. For example, the Many Labs Replication Project found only roughly 36% (Klein et al., 2014) and 54% (Klein et al., 2018) of studies to replicate respectively—leading to the proclamation of a 'replication crisis'.[1]

---

[1] Adjacent fields such as experimental philosophy fared better at around 70% (Cova et al., 2018), while an early replication project in economics replicated 61% of studies (Camerer et al., 2016). These failures

Another crisis that has been identified is the 'theory crisis', that various social sciences face a "lack of a cumulative theoretical framework" (Muthukrishna & Henrich, 2019, p. 221) and that theories not just methods have shaky foundations leading to failure of generalisation and replicability more broadly (Eronen and Bringmann, 2021; Fiedler, 2017; cf. Fried, 2020; Maatman, 2021, Others have proclaimed an 'applicability crisis', which is motivated by the claim that scientific findings are not as readily applicable as the scientific literature suggests. For example, when findings from the 'nudge' literature have been applied in large-scale contexts, they often failed to replicate or replicated only at a substantially reduced effect size (see e.g., Della Vigna & Linos, 2020). Based on these findings, some have argued that the social sciences as a whole are not (yet) in a position to give confident actionable advice and are thus in an applicability crisis.[2]

Several underlying causes of and potential solutions to these crises have been identified. One such cause is publication bias (Renkewitz & Keiner, 2019), that statistically significant results are the deciding factor for publication (Franco, Malhotra, & Simonovits, 2014). A further cause is the prevalence of questionable research practices, or QRPs (Fiedler & Schwarz, 2016), which are scientifically misguided but socially acceptable practices that compromise the integrity of scientific conduct. Other causes are selective analysis of some variables, dropping of experimental conditions, additional data collection after data analysis, warped incentives, and bad statistical (Gigerenzer, 2018) or measurement training (Lilienfeld & Stroher 2020).

Several potential solutions have been proposed under the banner of 'Open Science'. For instance, some have argued that preregistration can provide a strong counterbalancing force by making QRP's harder to execute and thus forcing researchers to adhere to pre-stated statistical analyses (Nosek et al., 2018). Others have proposed a new submission format of 'registered reports', in which only the hypotheses and the design of the study are subjected to peer-review and, if accepted for publication, result in a guaranteed publication irrespective of the findings. This solution has already been adopted by several journals (e.g., Chambers, 2013; Eder & Frings, 2021; Hardwicke & Ioannidis, 2018; Keil et al., 2020). Others have suggested better statistical education and reform in social science departments to remedy the statistical and methodological causes of the crises (e.g., Gigerenzer, 2018; Lakens, 2019).

In this paper, we want to draw attention to another potential partial cause of these interlinking crises and propose a solution to it that has not yet been picked up. Specifically, we argue that the structure of academic papers contributes significantly to the current situation of the social sciences, and that eliminating the discussion section promises to substantially contribute towards allaying at least some of these problems. Importantly, we see this solution as working in tandem with the other science reform

---

Footnote 1 continued

to replicate have been detected in many other domains such as mortality salience (Klein et al., 2019), metaphorical priming (Shanks et al., 2013), or ego depletion (Hagger et al., 2016).

[2] This applicability crisis has especially come to the fore during the COVID-19 pandemic. Social scientists argued that social science research ought to structure the policy response, e.g., concerning behavioural considerations of vaccine uptake (Kowal et al., 2020; van Bavel et al., 2020, p. 460; WHO, 2020), while IJzerman concludes, however, that social science is not yet "mature enough" (IJzerman et al. 2020, p. 1094) to offer advice to policy maker.For further work on the applicability crisis, see Loyka et al. (2020) and Dennis (2013).

efforts. Furthermore, we do not claim that our proposal is without drawbacks, and we offer a comparative argument by showing that its benefits outweigh its costs, and that it might meaningfully contribute to reforming social science. Before we move on to our discussion of this proposal in Sect. 4, we want to state the importance of this project and focus on two central challenges facing social science research in Sect. 3 which then will be addressed by our suggested reform.

3 Why this crisis matters?

The first aim of this section is to outline the scientific and societal consequences of these crisis. Our ultimate aim is to set up two challenges for conducting social science research in the current social structure of publication and public communication. We divide this section into (Sect. 3.1) concerns about public communication and trust in science and (Sect. 3.2) concerns about achieving the epistemic aims of science, specifically concerning the incentive structure of social science research.

3.1 Science communication: trust in social sciences

In analysing why these crises matter, we are focusing first on effective science communication and public trust. Trust in science matters because, as Wilholt (2013) states, "[p]olicy-makers, legislators, investors, and activists, as well as 'ordinary people' in their capacities as citizens or consumers frequently rely on the results of science, trusting […] these will help them make well-informed decisions" (Wilholt, 2013, p. 234). One should differentiate science-to-science communication from science communication to the general lay public.[3]

   We want to start with the latter. Especially the replication crisis has directly entered the public discourse, including ample media coverage[4]. As Fetterman and Sassenberg (2015) contend, the replication crisis is bound to have negative reputational effects on science. Recently Hendriks et al. (2020) showed that study credibility and researcher trustworthiness increase significantly if a study was successfully replicated and decreases otherwise (cf. also Mede et al., 2020). Wingen, Berkessel & Englich (2019) showed that low replicability specifically reduces trust in psychology. As such, the replication crisis already directly impacts science communication and consequently becomes important for scientific testimony (Gerken, 2015, 2020). This is because, as Whyte and Crease (2010) argue, one important project for a philosophy of science is to "facilitate trust between scientific experts and ordinary citizens" (cf. also Irzik and Kurtulmus, 2019; Whyte & Crease, 2010, p. 411) This project seems especially relevant against the background of wide-spread denial of various scientific

---

[3] And potentially also separate it from science communication to public policy makers since they have their own political aims and frequently some level of expertise that is not expected of a lay audience.

[4] These include: https://www.nytimes.com/2018/11/19/science/science-research-fraud-reproducibility.html, https://www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/. Veritasium published a YouTube video titled "Is Most Published Research Wrong?" that reached 2.7 million views: https://www.youtube.com/watch?v=42QuXLucH3Q.

findings.[5] However, the crises directly indicate that some of the present public distrust might be warranted, making public communication increasingly challenging by making differentiation of the levels of credibility of findings difficult. Thus, the crises impose a serious challenge to science communication as not to lose public trust in science overall, which would by itself lead to further negative outcomes such as a failure to comply with public health messaging.

Second, this problem of trust reaches beyond public science communication. It likewise concerns science-to-science communication, and as such directly impacts epistemic matters. As Romero (2019) argues, a social science in crisis can lead interdisciplinary research astray, which is especially troublesome for philosophy, since "empirically informed philosophers, and specifically moral psychologists, have relied heavily on findings from social psychology. They also need to clean up their act" (Romero, 2019, p. 7).

A further area in which the crises discussed here impact science-to-science communication is within scientific collaborations. There are good reasons to think that in many scientific disciplines trust can be more epistemically basic than empirical evidence, indicating that decreasing trust could undermine scientific knowledge production, going against the very epistemic aims of science (cf. Hardwig, 1991). Specifically, in an age of 'Team Science' (Ledford, 2015), scientific research largely relies on collaborations, which in turn depend on trust in the scientific community. This is because individual researchers in a collaborative project often have only partial information and expertise in a specific area, making trust a crucial element for successful collaboration (Bird, 2010; Fricker, 2002; Gerken, 2015). As such, it has been argued (De Ridder, 2022) that the erosion of trust due to events such as the replication crisis and the discovery of a widespread use of questionable research practices, might cause an can impede effective collaborations within scientific teams, ultimately hindering the production of scientific knowledge.

3.2 Achieving the epistemic aims of science

The epistemic aims of science have been a topic of ongoing debate; the most prominent proposals include truth (Khalifa, 2022), knowledge (Williamson, 2002), and understanding (De Regt, 2017).[6] Some claim that the aim is objectivity or proclaim a value-free ideal of science. But even if such ideals are too ambitious, as Haack (2003) argues, explicitly and systematically aiming at reducing the biases of the individuals involved in knowledge producing processes is one of the main constitutive features of science. The consensus is that, at least in the long run, science should be error-correcting (cf. Laudan, 1981; Mayo, 2005; Peirce, 1958).

---

[5] For instance, polling of U.S. adults on climate change shows an outright denial or at least denial of its human cause from somewhere between 32% (Saad, 2017) to 52% (Funk & Kennedy, 2016) of the population.

[6] Various others, frequently related epistemic aims of science have been proposed, such as explanatory coherence (Lycan, 1988), predictive success (Forster, 2002; Forster & Sober, 1994), rational acceptability (Putnam, 1981), agreement/consensus (Kuhn, 1962), social-epistemic value (Khalifa, 2010) and empirical adequacy (van Fraassen, 1980).

The aforementioned crises point towards a concerning epistemic defect in our scientific methodology and as such to an obstruction in achieving the epistemic goals of science. We argue that one major factor for at least partially improving the state of social sciences comes from resolving a specific tension between the epistemic aims of science and the non-epistemic goals of individual scientists. We ultimately argue that the current structure of academic research papers is such that those aims are misaligned and that such a misalignment contributes at least in part to the crises facing the social sciences. We go on to claim that our proposed solution can, in tandem with other reforms, contribute to solving this structural problem. In what follows, we present the problem of *misalignment* in order to properly set up the remainder of the paper.

Already in the 1960s, Polanyi (1962) argued that scientific cooperation emerges to a large extent as an unintended consequence of the social structure of science. This idea became especially prominent in the *sociology of science* (Barnes & Bloor, 1996), building on the thesis that the social features of science emerge from self-interested actions of individual scientists. This bottom-up process of explaining scientific norms was picked up by various philosophers of science, such as Kitcher (1990) and Strevens (2011) in what was termed the *economics of science* (cf. also Stephan, 2012). The central claim is that it is beneficial for promoting the aims of science if the aims of individual scientists align with the epistemic aims of science, while a misalignment can cause epistemic malfunctions of various types and severities.

The most straightforward way to achieve alignment is by having individual scientists engage in their research practices for reasons that align with the epistemic aims of science anyway—scientists might conduct research because they themselves want to get at the truth (i.e., the epistemic aim of science). However, scientists are also (and sometimes primarily) motivated by non-epistemic aims, such as advancing their career, getting recognition, or receiving grants. We can call such personal non-epistemic aims *credit* (Boettke & O'Donnell, 2016, p. 11; Zollman, 2018).

In publishing, credit seems to be at least one primary aim. Today, high publishing frequency in high-ranking journals is the main currency of success on the academic job market and in the grant system across disciplines. Still, the same publishing practices that give scientists credit might after all align with the aims of science. For instance, if publishing credit is a reliable indicator for scientific performance and having higher performing scientists on higher positions is on average epistemically advantageous for the scientific enterprise, then the career goals of individual scientists and the aims of science align. That such an alignment is important for the emergence of successful scientific norms and is indeed very frequently in place, is one of the main arguments of both Kitcher (1990) and Strevens (2011).

However, there are numerous reasons to think that this alignment might not be as straightforward as sometimes assumed. We argue that the crises discussed above point towards a tension between the aims of science and the credit aim of scientists. Others, who recognize a similar conflict, such as Hackett (2005) and Sovacool (2008), argue that publishing practices overemphasize novelty. Heesen (2018) argues that the credit system in science publishing incentivises speed and impact at the cost of reproducibility, pointing directly to a connection with the replication crisis. Fidler and Wilcox (2018, Sect. 4.5) suggest in accord with Vazire (2018, p. 416) that the aim of protecting one's own professional reputation often motivates resistance to

the self-correcting nature of replication (cf. also Fetterman and Sassenberg, 2015). Plausibly, science has already developed strategies to align the non-epistemic aims of scientists with the epistemic aims of science.[7] This typical alignment strategy consists of [AM] an adversarial mode[8] of science research and [IS] linking it with the individual incentive structure.

We want to start by explaining *AM*. Scientists, like all humans, are plagued with blind spots, biases (Pashler & Wagenmakers, 2012), prejudices, fall prey to rationalization (Schwitzgebel & Ellis, 2017), are fooled by cognitive artefacts (Machery, 2017), or are faced with suboptimal incentives (Pashler & Wagenmakers, 2012). Thus, it is paramount to put scientific findings under a high level of scrutiny. The general idea is to balance the blind spots of one scientist with the knowledge of another, challenging the biases and prejudices of one scientist by others with different flaws. In general, this is done by scrutinizing the findings one scientist holds dear by others who have no incentives to prefer that particular method or theory.[9] In this paper, we essentially argue for an extension of the adversarial mode that is already deeply entrenched in science and scientific practice, and which also appears in various foundational theories of scientific methodology. For instance, it underlies very prominently the Popperian philosophy of science (Popper, 1934[10])—especially the idea that we want to try our best to falsify scientific hypothesis under very rigorous conditions.[11]

A prime example for the adversarial mode of science is peer-review. Ideally, in peer-review, papers are put under high scrutiny by anonymous experts in the respective domain and only those papers who survive the most stringent of reviews will be accepted or invited for revision. However, the challenge for the adversarial mode in peer-review and elsewhere is this: Why would anyone ever want to put themselves under such scrutiny, especially if one's career depends on it? We now want to introduce the second part of the alignment strategy, i.e., linking the adversarial mode with the incentive structure of individual scientists. Consider peer-review again. Today, it usually brings with it much higher reputation to publish peer-reviewed than non-peer reviewed articles (Csiszar, 2016), and it typically brings with it a substantially

---

[7] A similar idea is present in Strevens (2011, p. 6) who views the alignment of norms and individual motivation as one of the three important dimensions of how the social norms of science and individual interests relate.

[8] A different, complimentary, strategy which we will not address here is *independent confirmation*. As Hacking (1983) points out, independent confirmation is one of the main contributors in differentiating valid results from mere artefacts of the testing apparatus or testing methodology.

[9] Some philosophers (cf. Moulton, 1993) argue that an adversarial discussion style might create a hostile environment. Note, however, that the focus of such arguments is usually on the discussion-style in research presentation and less on an adversarial approach to testing and reviewing. Also note that an adversarial mode does not have to be interpreted as mistrust, cf. also Nosek et al. (2012, p. 626).

[10] Importantly, they explicitly cite replication as a quality criterion of science.

[11] As an example for how the adversarial approach is explicitly put into practice see, for instance, Fischer et al. (2020) in experimental philosophy, where the members of a joint project are chosen because of their disagreements on the philosophical views in question. See also Clark and Tetlock (2022) for a proposal of 'adversarial collaborations' in which authors of different ideological backgrounds collaborate. While this is not what we see as the main challenge for social science research, it provides an additional example of this type of solution.

higher reputation to publish in journal with very stringent standards than with moderate standards.[12] This can be inferred from the fact that, nowadays, almost all current high-ranking journals list their (high) rejection-rate as a quality criterion which shows a perceived link between an adversarial mode and quality. On this basis, some countries have official rankings of journals which are drawn upon in hiring and promotion circumstances, as is the case in Finland where the Publication Forum classification system explicitly ranks journals.[13] As such, researchers who put their work under stronger adversarial scrutiny and succeed obtain more credit, and thus the self-interested scientists' career-goal and the epistemic aims of science align (at least, if the system works as intended).

This concludes our discussion of how the alignment strategy should work in practice. In this section, we argued that the epistemic aims of science rely on bias reduction and self-correction—features deeply associated with study replication. The sociological analysis of the incentive structure of science suggests that epistemic achievements of science can be diminished if the self-interested career-goal of scientists and the epistemic aims of science are misaligned. In this context, we presented a typical alignment strategy consisting in an adversarial mode of science research and linking it to the individual incentive structure. In the next section, we show how the current structure of scientific papers works against the presented alignment strategy, how this results in contributing to the crises of social science, and how this problem might be resolved in an as of yet unexplored way.

4 A further cause and a proposed solution: the elimination of the discussion section

In this section, we first (Sect. 4.1) identify an unexplored potential contributor to the interlocking crises facing social science: the structure of academic articles. We claim that the way researchers (are expected to) structure their research articles might set them up to engage in behaviours that feed into the crises and exacerbate other epistemic defects: Solving this necessitates structural changes. In Sect. 4.2, we then go on to propose a potential structural solution to this cause and argue that research in the social sciences might benefit from an elimination of the discussion section in papers. Arguing based on the alignments strategy as discussed in Sect. 3, we claim that this holds primarily because it reduces biases and possibilities to portray one's own research in a favourable but incorrect light, and further sets up an incentive structure for researchers to critically examine research of others via the novel proposed vehicle of a discussion paper, both of which jointly promise to contribute towards addressing the state of crisis the social sciences find themselves in. In Sect. 4.3, we discuss potential objections to this project and conclude with a summary of the costs and benefits of this approach.

---

[12] For an analysis of the merit of peer review, see Heesen and Bright (2019) and Schroter et al. (2008).

[13] https://julkaisufoorumi.fi/en/evaluations/classification-criteria.

4.1 The epistemic faults of a discussion section

Academic articles in the (social) sciences have roughly four main sections in addition to an abstract and a conclusion: (i) introduction and literature, (ii) methods, (iii) results, and (iv) discussion. The introduction sets up a problem, motivates the hypotheses, and contextualises the research. The methods section states the procedures, the sample selection process, and all further design implementation steps. The results section summarises all results and reports them in tables, graphs, and written form as well as though additional numerical in-text descriptions. In the discussion section, the results are put into context and conjectures as to the generalisability, limitations, and applicability of the findings are laid out (Bazerman, 2004, pp. 207–208). Specifically, it is in the discussion section that researchers provide verbal interpretations of their data by summarising the main findings and drawing attention to what they take to be the central take-away. They then frequently state limitations of both statistical and methodological nature and provide caveats to both these limitations and the findings presented in the paper. Depending on the discipline, this is followed by a rough outline for the practical applications of these findings by individuals, governments, or institutions.

The focus of our proposal is the discussion section. Historically, the discussion section in this modern form has been a consistent and recognisable part of academic papers for around 100 years. Atkinson shows that while historically, experimental papers during the 17th and early 18th century were largely "unelaborated, miscellaneously organized, and relatively narrative in character" (Atkinson, 1998, xxiv), by 1775, some predecessors to a discussion section were already present in recognisably similar form. By the 19th century, the rough structure of a theoretical part followed by the experiment followed by a discussion was relatively common, and by 1925 it had become "the standard" it is now (Atkinson 1998, 70; cf. also Bazerman, 1985).

We claim that this structure of academic articles carries with it several epistemic flaws that have prohibited science from functioning as well as it otherwise might have. Further, we argue that this structure may also have directly and indirectly contributed to the set of crises. Specifically, we claim that discussion sections directly foster behaviours that rest upon epistemically dubious grounds, such as enabling researchers to set the narrative of their results, allowing them to put the focus on certain results, and enabling them to self-report the limitations they see in their own design. These behaviours all fall prey to cognitive biases such as the choice-supportive bias, post-hoc rationalization, ostrich effect, bias blind spot, or the hindsight bias. Additionally, this system is perhaps best characterised as consisting of several perverse incentive structures. As such, researchers are less likely to honestly report the data, their resultant true implications, and the applicable methodological drawbacks. This current situation runs contrary to the alignment strategy as presented in (II), since for these behaviours to be in accord with the epistemic aims of science, all researchers would have to be immune to these biases of various types and would further have to exhibit an unreasonably high degree of selflessness. This, we claim, is unlikely.[14]

---

[14] For instance, the results presented by Anderson et al. (2010) suggest that current incentive structures of science are hard to overcome for individual scientists even if scientists have the epistemic aims of science in mind.

Let us discuss those shortcomings one by one. First, in the discussion section, researchers can put into focus easy-to-explain data that fit their narrative while dropping entirely the data that do not fit or that are even counter to the proposed narrative. This behaviour is often a clear ethical violation of research conduct (cf. Greenwald, 1975). However, it is yet incentivized by standard publishing practices that reward presenting to the editor and reviewers a paper with a clear narrative that neatly fits all the data rather than a paper where the overall story is less clean but closer to the actual results. Doing the latter reduces the prima facie chances of publication: This is why researchers often use the discussion section to draw attention to the data that do fit their narrative and the hypotheses reported, while sweeping those parts that would make the paper less convincing under the rug (or into the appendix, which is sometimes located online behind several steps to ensure that it cannot be easily accessed, if at all). This is a clear misalignment of individual self-interested incentives and the aims of science. It results in a net negative impacting on science as it is, through directly promoting selective reporting. It also negatively impacts science communication more generally as discussion sections fail to be accurate representations and explanations of the data collected.

One might object that this is only a minor problem since scientists are aware that the discussion section is bound to epistemic defects. It is probably true that an increasing number of scientists are becoming aware of the issue, however, not all readers in the academic sciences possess the requisite level of proficiency to assess the specialized data, and this is especially difficult for science reporters and in collaborative and interdisciplinary research. Generally, even for scientists who are well aware of the potential for bias in discussion sections, it can pose major difficulties to separate out the true findings from the noise, particularly if the discussion section is very persuasive or engaging. Therefore, switching to a less biased system is preferable to relying on the awareness of the readers.

Second, researchers are also asked to state the limitations of their research design in the discussion section. This, while on the face useful as the authors are plausibly best positioned to identify where the weaknesses of their research design are and which corners have been cut, is equally problematic because this again leaves it up to the discretion of the researchers themselves to point out limitations. One need not come up with far-fetched scenarios to imagine researchers downplaying the trade-offs that they had to take in their experimental design and the resultant limitations of the results. On top of that, these sections are also often accompanied by a short explanation for why these limitations do not fully apply to the design reported and why they ought not be taken as too impeding (to both the publishability of the research and the widespread adoption of the finding). Researchers may then be more likely to not honestly state the full extent of the limitations, either because they themselves are suffering from cognitive biases that make it hard to see their work in an objective light, or, more likely, because they have incentives not to do so as they seek publication as well as public praise and recognition. Though this practice can sometimes be rectified by the peer-review system through its adversarial element, we claim that often, researchers understate the limitations of their own research to the detriment of science and public trust more generally in a way that is hard to evaluate from a sporadic peer reviewer's perspective. Moreover, this problem is made worse by university press offices, that

often continually overstate the findings reported and discussed in discussion section to an even greater extent than the authors themselves, making the problem even worse by extending it directly into the science-to-public communication front.

To combat these shortcomings of the discussion section, there are guidelines from journals and professional societies that outline best practices for what goes into a discussion section and how to properly engage with one's own data. The concern is, however, how to incentivize researchers to follow such guidelines and to effectively self-police. Both worries discussed above have in common that for science to function properly, researchers would have to act against their career-guiding publishing aims by honestly discussing non-conforming data and by openly stating the true limitations of the design, making it increasingly unlikely that this is indeed happening at a large scale. As such, it runs contrary to our proposed alignment strategy - missing the adversarial pillar. Changing dishonest or biased behaviour is unlikely to come about without addressing the underlying defective misalignments of individual self-interested aims and the aims of science and cannot as such be laid at the feet of individual researcher's responsibility, but rather must be solved systemically and in tandem with existing reform efforts.

4.2 Removing the discussion section

These shortcomings have not gone undetected (e.g., Barbour, 2015; Edwards and Roy, 2017). However, we proffer a novel solution: the wholesale elimination of the discussion section from academic papers. This brings with it not only a (partial) redress of the original problems outlined in Sect. 4.1, but might also lead to various additional theoretical and practical upsides that themselves impact the final cost-benefit analysis of this proposal. We see this structural change as working alongside other science reform efforts and not as a standalone solution; in fact, it may lack much of its potential benefits if other aspects of the scientific process are not improved upon. To begin with, let us consider how eliminating the discussion section might promise to address the outlined challenges.

First, removing the discussion section directly addresses the problems of researcher incentives in relation to the discussion of non-confirming data and serious limitations as there is simply no more discussion section to do this in. In our model, research articles introduce a problem, state their (preferably pre-registered) hypotheses clearly (Introduction), present the design (Methods), and report their (preferable analysis-plan based) findings (Results). In such a model, researchers are no longer able to selectively discuss their data or limitations in the discussion section, and it would be significantly more difficult to have these sentiments appear in other sections of the paper. Interacting with other approaches and solutions, such as pre-registration or analysis plans, researchers would be further incentivized to outline all their data as stated in the pre-registration/registered report. Without having a place in each research paper in which researchers are heavily incentivized to misrepresent their contributions, understate their limitations, and overplay their practical importance, there is significant reason to believe that the misalignment is, at least to a significant extent reduced. We argue that doing so will not only be better for science but might also be preferred by researchers

as they are then able to conduct their scientific work more straightforwardly, with less of an incentive to oversell their results, thus reducing inner personal conflicts where present. Note that this move is more akin to reducing the opportunities to do harm and thus indirectly reducing the incentives.

That being said, any structural changes of this magnitude will have unforeseeable consequences. One risk might be that researchers, no longer being able to frame their results as they please in their discussion sections, will simply move to misstating their data. While we cannot rule this out, our argument for this proposal does state that it can only be expected to have substantial positive impacts if it is implemented in conjunction with other reforms like pre-registered analysis plans. Due to the fact that Open Science initiatives, such as pre-registered analysis, specifically target questionable research practices related to statistical analysis within the data section, it is a much more difficult task to shift bias towards this section than expressing the bias in the discussion section. While researchers may resort to extreme measures such as manipulating the data itself, if they are willing to go to such lengths, it is unlikely that they are not already doing so within the current system. As such, while there is a potential that researchers move their bias from the discussion section to the data section, we at least can be confident in the minimal claim that without the discussion section, the misalignment will be reduced (though is unlikely to be eliminated).

Additionally, we claim that most of the upside of our proposal will be cashed out by our second proposal of a novel type of academic article as a replacement of the discussion sections: a discussion paper.[15] Discussion papers are papers designed to discuss one or more original research articles (or the data presented within them). They are aimed at contextualising the findings, outline future research questions, and analyse limitations that allow careful interpretation of the results and appropriate practical guidance. Contrary to the current format, where only researchers themselves write the discussion section, discussion papers can be written by a different set of researchers (that may or may not include the authors of the paper reporting the data), thus directly drawing on the better epistemic ability of researchers to evaluate others' work in an unbiased way. Specifically, having discussion papers written by somebody other than the researchers has the epistemic upsides of resulting in (i) personal bias reduction, (ii) a utilisation of the division of academic labour (potentially across disciplinary boundaries), (iii) an introduction of novel incentives in line with the adversarial mode of scientific research, and (iv) an improvement of science communication downstream.[16] Let us tackle those upsides in turn.

(i) Outsourcing the discussion section to papers not written by the authors of the original papers plausibly reduces personal biases across the board by using our proposed alignment strategy. This is because the authors of the discussion section do not

---

[15] Though there may be other solutions that emerge from a removal of the discussion section, and we believe that eliminating the discussion section is potentially worth it irrespective of the structure that emerges as a result.

[16] Some of these upsides would also be present if the discussion paper was indeed written by the same authors (perhaps because they know their complex data set very well): This may still reduce personal biases (i) as the publication success of the reported data does not depend on the discussion section at all, and further collaboration (be it adversarial or not, cf. Heyman et al., 2020) may also lead to some of the other advantages (ii)-(iv).

share the same incentive structure and personal involvement with the original research. They are in a less biased position to evaluate how the data fit into the bigger picture and what the actual limitations are. Outsourcing the discussion section thus allays some of those worries and promises an incentive structure that is less likely to coincide with cognitive biases to produce subpar scientific results. In simple terms, it helps the scientists align their personal incentives with the goals of science. This is plausibly even the case when the authors of the discussion section are the authors of the original empirical work since the publication of their data no longer depends on their framing of them in the discussion section, at least partially reducing the personal biases in play.

There is a potential risk that separating the discussion section from the main article may amplify the problems associated with an integrated discussion section. It might be argued that authors now have more incentive to make the discussion paper attention-grabbing, leading to a misalignment of personal aims and the epistemic aims of science. However, we believe that the peer-review process can help mitigate this risk. In the current system, a biased discussion section may receive less scrutiny since the referees have to divide their focus and may prioritize getting research published simply because the data is important even if the quality of the discussion section is subpar. In contrast, in a system with a separate discussion paper, all scrutiny of the referees is focused on the quality of the discussion paper alone. Therefore, while we cannot completely rule out the possibility of a new set of bad incentives, we argue that the peer-review process is now in a better position to ensure the integrity of the discussion paper.(ii) Splitting off the discussion section from the primary data papers also allows academic research to directly harness the fruits of the division of labour in a majority of cases. Specifically, our proposal might result in an altered research landscape where those who are especially apt at designing and conducting studies do so, while those with a more generalist skill set synthesise several such results into discussion papers, perhaps along an experimentalist-theorist divide that is already seen in other disciplines. This advantage is especially striking against the backdrop of Cohen's (1990) observation that in psychology, researchers frequently misinterpret p-values and Ziliak & McCloskey's (2008) contention that empirical researchers generally too often draw wrong conclusions regarding the statistical significance of their results. Dividing the labour between scientists who are specialised in conducting studies and scientists who are specialized in interpretation and synthesis promises an improvement that might help address the crises facing the social sciences while also plausibly boosting scientific productivity. Additionally, this division of labour may even be beneficial across disciplines. For example, consider a philosopher writing up a discussion paper of research on moral judgements in addition to a similar paper being written by a psychologist. Both types of discussion papers would bring a different skill set to bear on the available data which may then allow readers to get a perspective on the data that would otherwise be inaccessible. Additionally, cross-disciplinary division of labour within discussion papers might be especially useful, where authors from different disciplinary backgrounds collaborate on a single discussion paper, drawing on research from several disciplines to allow for a more balanced and holistic picture of scientific research in a given area of study. This, so we argue, would greatly improve scientific progress within, but also between disciplines.

(iii) We take our proposal to improve upon the incentive structure to better align the credit incentives and the epistemic aims of science. This contrasts with the previous model in which authors were incentivised to portray their studies as without serious limitations and as providing actionable recommendations for policy makers. The revised model improves this by reducing perverse incentives, leaving authors with less reason and less opportunities to act contra the aims of science, at least concerning interpretation, limitations, and applications of the research. Further, having an adversarial relationship between those writing the discussion papers and those writing the data focused papers does not only eliminate or reduce the worry of bias and perverse incentives, it also independently improves the epistemic environment of researchers by removing epistemically unfavourable elements that make researchers prone to self-deception (cf. Heyman et al., 2020). This then directly interlinks our suggestion with previously proposed solutions to the crises facing the social sciences. Since the peer-review process will be focused entirely on the design and results for research papers and entirely on the merit of the discussion for discussion papers, this incentivises authors of experimental papers to pre-register their data and make their data sets open accessible, as these practices are now directly conducive to publication success, interlinking this proposal directly with other Open Science reform efforts. Further, the novel discussion papers themselves will have a distinct incentive landscape, in which authors might be more likely to discuss limitations and applications of research more honestly as portraying these data in a good light is at least not central to their success in publishing.

(iv) Finally, this solution promises to improve public communication of scientific findings. Usually, public science reporting draws on the interpretations of scientific studies by making them easier digestible for a general audience. After all, page-long regression tables are often not what can be communicated to the public. However, since the incentive structure of the standard discussion section motivates overstatement and distortion of the empirical findings, these defects get passed on directly to the public. This not only leads to potential misinformation, but if some of those overstatements are recognized, it may also lead to an increase in general mistrust in scientific findings. Since our proposal reduces problematic incentives, and thus promises to decrease overstatement and distortion of findings, it promises to improve public communication of scientific findings as a consequence. Of course, it may be that science journalists will not engage with the discussion papers meaningfully; after all, they are already likely to skip the discussion section. We argue that while this is true, the mere existence of and potential institutional prestige of high-quality discussion papers may also make it more likely to journalists to pick them up and report on them. This may then take less the form of a sensationalist piece on a specific finding, but rather more general summaries of the state of the scientific field. Moreover, these discussion papers will still allow journalists to pick out specific findings that translate well into a headline, but the whole paper itself will arguably be more likely to contribute to a more balanced and nuanced depiction of the science.

Given these advantages of splitting research papers into two parts, this should give academic journals some incentive to switch to such a mode. As argued in (Sect. 3.2), an adversarial mode is already valued in current publishing to increase journal reputation. Thus, journal publishers could benefit from seriously considering our proposal to further increase their reputation, making this proposal also plausibly implementable in the short term. Here is what we specifically propose: Journals ought to disallow the use of a discussion section (and its contents) in their primary research articles which would only be consisting of an introduction, a methods section, a results section, and a conclusion. They would then also start to accept manuscripts of 'discussion papers'. These discussion papers discuss the data of one or multiple primary research papers. Authors of these research papers would then be asked to critically discuss the data by pointing out limitations of the designs, highlighting potential applications, drawing out interesting follow-up opportunities, and synthesising the results in the wider literature. This is markedly different from the current system and a radical change, but one that promises to contribute to improving social science alongside other reform efforts.

4.3 Objections and concluding remarks

We close this section by considering several objections before giving a brief comparative argument of the benefits and costs of our proposal. First, one might argue that the discussion sections themselves are not the fundamental cause of any of these crises. This, so the objector, is because what is truly driving the challenge is the overall incentive structure of science, and as such, the overall incentive structure ought to be identified as the cause and be addressed directly. On this line of reasoning, proposing to eliminate the discussion section would be akin to merely treating a symptom and thus failing to actually address the root cause. While this assessment might be partly correct, its conclusion does not follow. From the claim that the problems with the discussion section ultimately stem from the overall incentive structure facing scientists it does not follow that eliminating the discussion section is misguided. Rather, addressing one aspect of this incentive structure related to the discussion section might go a long way towards affecting the overall structure in reverse.

As a second objection, one may object that the above approach relies on other solutions already being implemented successfully. Specifically, the proposal to publish research papers primarily consisting of methods and results might presuppose that data are being shared openly and freely according to Open Science best practices, as failing to disclose data open and freely merely shifts the problem of individual bias. We agree that eliminating the discussion section alone cannot solve the multitude of crises facing the social sciences and that it will most likely require a multiplicity of different reforms to have an effect. However, we do not see this as a problem as no single solution is capable of fixing the interlinking crises and argue that eliminating the discussion sections can play one part in addressing them.

Third, a downside of outsourcing the discussion section might be that some data of research papers might never be discussed. This can be frustrating for the researchers who published the research paper and want engagement with their findings. We reply that our model does not exclude the authors of the original research paper to also write

a discussion paper on their data, though their discussions would then go through a separate peer-review process that would evaluate them solely on their contribution to the discussion (not the data collection). Furthermore, discussion papers are stand-alone publications and as such attach themselves to the already existing incentive structure for writing papers in general (i.e., credit). This encourages researchers to engage with data that is not theirs but it also allows engagement with one's own data. It is, thus, plausible that we might even see more discussion of the data than we currently do. Moreover, authors of discussion sections have now an incentive to carefully study bulks of research papers, and editors and reviewers will additionally consider which one is worth discussing. This plausibly creates an environment that allocates the resources for discussion writing more efficiently than a system which rigidly limits itself to only one discussion of one research paper by the same researchers.

Fourth, one may object that the significant heterogeneity that is observed between disciplines with respect to replicability, questionable research practices, and the like may suggest that our framing (as well as our proposed solution) may not apply to all social sciences equally. In short, we think that this objection is roughly correct. It is the case that there are differences in replicability and publication of replications (e.g., Berry et al., 2017), which is something that forecasts continue to anticipate in the future (Gordon et al., 2020). As such, the elimination of discussion sections may have disparate effects across disciplines. However, we do not think that this is indeed an issue for our proposal as we do not think that discussion sections are the underlying cause of all the crises. Rather, as we have argued, we believe that removing discussion sections might have a net positive impact on some of the malaises facing some disciplines.

A fifth objection is that the proposed discussion papers create their own perverse incentives. It could be argued that the authors of the initial research paper would be very well positioned to immediately submit a discussion paper accompanying their first publication, effectively pre-empting further submission. This, so the objection, might lead to an unfair advantage on part of the original authors because they increase their likelihood at publishing more work in the limited journal space. We respond to this objection as follows: First, even if there is such an advantage for the original authors, in our proposed structure there is at least a reasonable chance for authors other than the original researchers to participate in a separate discussion of the data by publishing a discussion paper (perhaps by including additional data from other work). As such, our proposal might retain some perverse incentives, but they are arguably reduced. Second, if the previously sketched epistemic advantages of a division of labour hold, then at least some authors specializing in discussion papers will also have one advantage over the authors of the research papers who are not that specialized. As such, it is implausible that the authors of the original research papers will always be the ones who publish the discussion papers or will always end up the ones with the highest visibility, while it may also be the case that further, more expansive discussion papers discuss the data of any individual research paper, even if that paper's authors have already published a standalone discussion paper.

A sixth objection is that even if our proposal was implemented, there might be a dearth of potential authors willing to write such a discussion paper. It is unclear, so the objection, what will motivate the work on discussion papers such that the academic

credit system will incentivise authors to spend substantial effort towards writing these papers. We respond to this in two ways: First, as we have argued above, it is very likely that the majority of discussion papers will include authors that have authored (some) of the underlying experimental papers, thus providing intrinsic motivation to write them. Second, and more importantly, we think that in situations where authors do not write the discussion papers themselves, career incentives to publish in prestigious journals will most likely provide sufficient incentives for third-party writers to write these papers. We believe that such high-profile venues will exist just by the nature of there being venues that offer discussion papers, and some of these being ranked higher (on whatever metric). Polemically put, we have yet to see an academic system (be it universities, journals, etc.) that academics themselves have not turned into a prestige hierarchy, accompanied by a drive to be on top. We expect that this mechanism will extend to our proposal as well, incentivising the production of high-quality discussion papers. Furthermore, we add the following anecdote: The journal *Data* is a data-science journal that offers (among other options) two types of submissions: Data Descriptors and Reviews. The former outline and explain a data set, the latter "concise and precise updates on the latest progress made in a given area of research". While this is not quite our proposal and it is not within the social sciences, we take this example of a proof of concept.

We have now outlined the benefits of our proposal and replied to objections. However, as most policy prescriptions, our proposal is not without potential drawbacks. After having spent significant effort above outlining the potential upsides of our proposal, there are several costs to consider. First, one central cost with our proposal is that we may lose the epistemic advantages of authors discussing their own data in some instances. Particularly, the authors of a study often have unique insights into their data that may not be immediately apparent to third-party researchers. This is especially true for studies that involve complex datasets. By outsourcing the discussion section to third-party authors, we may miss out on important nuances and insights that only the original authors can provide. Furthermore, the original authors may have access to additional data or information that is not included in the published study. This may include preliminary analyses, unpublished data, or insights gained through personal experience or interactions with study participants. Keep, however, in mind the, as we argued, significant epistemic cost associated with the traditional structure of academic papers. Therefore, it is important to weigh the potential costs and benefits of our proposal in a nuanced and comprehensive manner. Note further that this only partly applies since the original authors can still write discussion papers as well. It is expected, however, that on various occasions original authors will never write or be able to publish discussion papers relating to their research paper. In such cases, it is indeed a genuine cost of our proposal that stands in a trade-off with the epistemic advantages of our proposal, which we argue outweigh the costs substantially.

A second potential source of substantial cost that may be associated with our proposal is that implementing such a sweeping change is rife with uncertainties that are extremely difficult to resolve ex ante, and that this may lead to unintended consequences. However, this is a cost associated with most reform efforts, even though we cannot rule out backfire effects of one sort or another. Additionally, even if our proposal is met with acclaim, then it is to be expected that initially only a few journals

will pick up on the idea, as has been the case with reforms such as registered reports (Chambers, 2013; Hardwicke & Ioannidis, 2018; Keil, Gatzke-Kopp, Horvath, Jennings, & Fabiani 2020). As such, the whole system is only reformed gradually and potential drawbacks, once properly identified, can be met along the way. Even if the trade-offs stay initially incommensurable, having a diversification of journals between those with our newly proposed structure and other which follow a more traditional publishing structure will over time get us a better outlook on the benefits and costs of our proposal and the current system.

Overall, do the benefits outweigh the costs? Throughout (Sect. 4), we have outlined several potential benefits as well as costs associated with our proposal. In short, our proposed structural reform might lead to substantial improvements in bias reduction efforts, better incentives, and scientific integrity overall. However, there are costs to this that have to be weighed against the benefits. such as we might lose the epistemic advantage of authors discussing their own data in some instance. A general risk of such a major reform is that it may lead to a new equilibrium of behaviours that end up having even worse incentive structures building on a whole new set of biases. As we have said before, we cannot be sure that this does not happen. However, we believe that, because we see our proposal as a proposal on top of already existing reform movements, and because any type of adaption will be gradual, that most of the costs may be manageable. Furthermore, if we consider such a maximally pessimistic forecast, we should also consider that the potential maximum value created by this proposal is extremely high, promising widespread improvements across the social sciences.

Lastly, let us return to the original question of the crises the social sciences are facing and the potential progress that elimination of discussion sections may make. We argue that eliminating the discussion section might contribute meaningfully to addressing these crises, though, as pointed out before, much of this has to happen alongside other reform efforts, and any such effort is not without costs. Having removed the discussion section from papers, research norms will have to shift regarding data presentation and availability, potentially in line with Open Science norms. Doing so might directly impact replicability concerns, as authors are now preparing their data sets for discussion papers (if they want their research to be included in future papers), thus aligning their incentives with those that would contribute to combating the replication crisis. Further, dedicated discussion papers would mean authors have no incentive to play up a study's relation to theory. This will allow readers to be in an epistemically superior position to judge which empirical investigations ought to inform scientific theorising and which fail to meet these standards. This is again a marked improvement over the previous system that might contribute to addressing the theory crisis and by moving social science towards its goal of being a cumulative science in line with Muthukrishna and Henrich (2019, p. 221) who argue that the theory crisis stems from a "lack of a cumulative theoretical framework". Our proposal would make accumulating several empirical findings and investigating them at once holistically easier, making it more straightforward for further researchers to directly build upon it. Finally, removing the discussion section also removes most of the place in which authors can state the wide applicability of their research that, as argued above, is frequently overstated. Discussion paper authors would face a different set of incentives and would, as such, be more

likely to accurately portray how some data could impact public policy or adaption in the private sector, and additionally would also be evaluated separately in peer-review, thus providing a plausible path to combating the applicability crisis. Finally, by combating overstatements of applicability and selective interpretations of findings, this will additionally lead to a more honest public communication of scientific findings, promising to reduce mistrust in science. As such, our proposal, in tandem with other reform efforts, promises to be one important contribution in addressing both concerns of the scientific crises and of science communication. We hope that our paper can start a debate on this (or similar) proposals.

Declarations

Competing interest The authors report no competing interests regarding this work and its funding.

References

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2010). Extending the mertonian norms: Scientists' subion to norms of research. *The Journal of Higher Education, 81*(3), 366–393.

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, *3*(3), 266–286.

Atkinson, D. (1998). *Scientific discourse in sociohistorical context: The philosophical transactions of the royal society of London* (pp. 1675–1975). Routledge.

Barbour, V. (2015). Perverse incentives and perverse publishing practices. *Science Bulletin*, *60*, 1225–1226.

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.

Barnes, B. D., & Bloor, J. H. (1996). *Scientific knowledge: A sociological analysis*. University of Chicago Press.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written communication*, *2*(1), 3–23.

Bazerman, C. (2004). *Communicating science: The scientific article from the seventeenth century to the present*. Oxford University Press.

Benessia, A., Funtowicz, S., Giampietro, M., Pereira, Â. G., Ravetz, J. R., Saltelli, A., & van der Sluijs J. P. (2016). *Science on the Verge.* Consortium for Science, Policy & Outcomes. Tempe, AZ and Washington, DC

Berry, J., Coffman, L. C., Hanley, D., Gihleb, R., & Wilson, A. J. (2017). Assessing the rate of replication in economics. *American Economic Review*, *107*(5), 27–31.

Bird, A. (2010). Social knowing: The social sense of 'scientific knowledge'. *Philosophical Perspectives*, *24*, 23–56.

Boettke, P. J., & O'Donnell, K. W. (2016). The social responsibility of economists. In G. DeMartino & D. McCloskey (Eds.), *The oxford handbook of professional economic ethics* (pp. 116–136). Oxford University Press.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at cortex. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 49*(3), 609–610.

Clark, C. J., & Tetlock, P. E. (2022). Adversarial collaboration: The next science reform. In C. L. Frisby, R. E. Redding, W. T. O'Donohue, & S. O. Lilienfeld (Eds.), *Political bias in psychology: Nature, Scope, and solutions.* Springer.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312.

Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., & Cushman, F. (2018). Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology, 12*, 1–36.

Csiszar, A. (2016). Peer review: Troubled from the start. *Nature*, *532*(7599), 306–308.

De Regt, H. W. (2017). *Understanding Scientific understanding*. OUP.

De Ridder, J. (2022). How to trust a scientist. *Studies in the History and Philosophy of Science*, *93*, 11–20.

Della Vigna, S., & Linos, E. (2020). *RCTs to scale: Comprehensive evidence from two nudge units*. Wiley.

Dennis, B. (2013). Validity crisis' in qualitative research. *Qualitative research: A reader in philosophy, core concepts, and practice* (pp. 3–37). Peter Lang.

Eder, A. B., & Frings, C. (2021). Registered report 2.0: The PCI RR initiative. *Experimental Psychology*. https://doi.org/10.1027/1618-3169/a000512

Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science, 34*(1), 51–61.

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788.

Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *Plos ONE*. https://doi.org/10.1371/journal.pone.0143723

Fidler, F., & Wilcox, J. (2018). Reproducibility of scientific results. In N. Edward Zalta (Ed.), *The stanford encyclopedia of philosophy (winter 2018 edition).* Stanford University.

Fiedler, K. (2017). What constitutes strong psychological science? the (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science, 12*(1), 46–61.

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, *7*(1), 45–52.

Fischer, E., Engelhardt, P., & Sytsma, J. (2020). Inappropriate stereotypical inferences? an adversarial collaboration in experimental ordinary language philosophy. *Synthese.* https://doi.org/10.1007/s11229-020-02708-x

Fletcher, S. C. (2021). The role of replication in psychological science. *European Journal for Philosophy of Science*, *11*(1), 1–19.

Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, *29*(2), 158–181.

Forster, M. R. (2002). Predictive accuracy as an achievable goal of science. *Proceedings of the Philosophy of Science Association*. https://doi.org/10.1086/341840

Forster, M. R., & Sober, E. (1994). How to tell when simpler, more unified, or less Ad Hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science, 45*, 1–35.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.

Fricker, E. (2002). Trusting others in the sciences: A priori or empirical warrant? *Studies in History and Philosophy of Science Part A*, *33*(2), 373–383.

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288.

Funk, C., & Kennedy, B. (2016). Public views on climate change and climate scientists. http://www.pewinternet.org/2016/10/04/public-views-on-climate-change-and-climate-scientists/

Gerken, M. (2015). The epistemic norms of intra-scientific testimony. *Philosophy of the Social Sciences*, *45*(6), 568–595.

Gerken, M. (2020). Public scientific testimony in the scientific image. *Studies in History and Philosophy of Science Part A*, *80*, 90–101.

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.

Gordon, M., Viganola, D., Bishop, M., Chen, Y., Dreber, A., Goldfedder, B., & Pfeiffer, T. (2020). *Are replication rates the same across academic fields? community forecasts from the DARPA SCORE programme*. Royal Society Open Science.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1.

Haack, S. (2003). *Defending Science—within reason: Between scientism and cynicism*. Prometheus Books.

Hackett, B. (2005). Essential tensions: Identity, control, and risk in research. *Social Studies of Science*, *35*(5), 787–826.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.

Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., & Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 546–573.

Hardwicke, T. E., & Ioannidis, J. P. (2018). Mapping the universe of registered reports. *Nature Human Behaviour*, *2*(11), 793–796.

Hardwig, J. (1991). The role of trust in knowledge. *Journal of Philosophy*, *88*(12), 693–708.

Heesen, R. (2018). Why the reward structure of Science makes reproducibility problems inevitable. *Journal of Philosophy*, *115*(12), 661–674.

Heesen, R., & Bright, L. K. (2019). Is peer review a good idea? *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axz029

Hendriks, F., Kienhues, D., & Bromme, R. (2016). Trust in science and the science of trust. *Trust and communication in a digitized world* (pp. 143–159). Springer.

Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research. *Public Understanding of Science*, *29*(3), 270–288.

Heyman, T., Moors, P., & Rabagliati, H. (2020). The benefits of adversarial collaboration for commentaries. *Nature Human Behaviour*, *4*(12), 1217.

IJzerman, H., et al. (2020) Use caution when applying behavioural science to policy. *Nature Human Behaviour, 4*, 1092–1094.

Irzik, G., & Kurtulmus, F. (2019). What is epistemic public trust in science? *The British Journal for the Philosophy of Science*, *70*(4), 1145–1166.

Keil, A., Gatzke-Kopp, L. M., Horváth, J., Jennings, J. R., & Fabiani, M. (2020). *A registered report format for psychophysiology*. Wiley.

Khalifa, K. (2010). Social constructivism and the aims of science. *Social Epistemology, 24*(1), 45–61.

Khalifa, K. (2022). Understanding, truth, and epistemic goals. *Philosophy of Science*. https://doi.org/10.1086/710545

Kowal, M., Bialek, M., & Groyecka-Bernard, A. (2020, December 11). Behavioural science is not good enough for building ROckets, but still useful in crisis. Retrieved from: https://socialsciences.nature.com/posts/behavioural-science-is-not-good-enough-for-building-rockets-but-still-useful-in-crisis?badge_id=569-nature-human-behaviour

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy*, *87*, 5–21.

Klein, R., Ratliff, K., Vianello, M., Adams, R., Jr., Bahník, S., Bernstein, M., & Cemalcilar, Z. (2014). Data from investigating variation in replicability: A "many labs" replication project. *Journal of Open Psychology Data*. https://doi.org/10.5334/jopd.ad

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B. Jr., Alper, S., & Batra, R. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490.

Klein, R. A., Cook, C. L., Ebersole, C. L., Vitiello, C., Nosek, B. A., Chartier, C. R., & Cromar, R. (2019). *Many labs 4 Failure to replicate mortality salience effect with and without original author involvement*. University of California Press.

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lakens, D. (2019). The practical alternative to the p-value is the correctly used p-value. *Perspective on Psychological Science*. https://doi.org/10.1177/1745691620958012

Laudan, L. (1981). *Science and hypothesis: Historical essays on scientific methodology*. Springer.

Ledford, H. (2015). Team science. *Nature*, *525*(7569), 308–311.

Lilienfeld, S. O., & Strother, A. N. (2020). Psychological measurement and the replication crisis: Four sacred cows. *Canadian Psychology/Psychologie canadienne*. https://doi.org/10.1037/cap0000236

Loyka, C. M., Ruscio, J., Edelblum, A. B., Hatch, L., Wetreich, B., & Zabel, A. (2020). Weighing people rather than food: A framework for examining external validity. *Perspectives on Psychological Science*, *15*(2), 483–496.

Lycan, W. G. (1988). *Judgment and Justification*. Cambridge University Press.

Maatman, F. O. (2021). Psychology's theory crisis, and why. formal modelling cannot solve it.

Machery, E. (2017). *Philosophy within its proper bounds*. OUP.

Mayo, D. (2005). Peircean induction and the error-correcting thesis. *Transactions of the Charles S Peirce Society, 41*, 299–319.

Mede, N. G., Schäfer, M. S., Ziegler, R., & Weißkopf, M. (2020). The "replication crisis" in the public eye: Germans' awareness and perceptions of the (ir) reproducibility of scientific research. Public understanding of science

Moulton, J. (1993). A paradigm of philosophy: The adversary method. In Hintikka (Ed.), *Discovering reality* (pp. 149–164). D. Reidel.

Muller, D. (2016). Is most published research wrong? *Veritasium*, https://www.youtube.com/watch?v=42QuXLucH3Q [05.03.2021].

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229.

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.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606.

Pashler, H., & Wagenmakers, E. (2012). Introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530.

Peirce, C. S. (1958). The logic of drawing history from ancient documents. In A. W. Burks (Ed.), *The collected papers of Charles Sanders Peirce* (Vol. IV, pp. 89–107). Belknap Press.

Polanyi, M. (1962). The republic of science. *Minerva*, *1*, 54–73.

Popper, K. (1934). *The logic of scientific discovery*. Routledge.

Putnam, H. (1981). *Reason, truth and history*. Cambridge University Press.

Renkewitz, F., & Keiner, M. (2019). How to detect publication bias in psychological research. *Zeitschrift für Psychologie*. https://doi.org/10.1027/2151-2604/a000386

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*. https://doi.org/10.1111/phc3.12633

Saad, L. (2017). Global warming concern at three-decade high in US. Retrieved October 14, 2017, from http://news.gallup.com/poll/206030/global-warming-concern-three-decade-high.aspx

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine*, *101*(10), 507–514.

Schwitzgebel, E., & Ellis, J. E. (2017). Rationalization in moral and philosophical thought. In J. F. Bonnefon & B. Tremoliere (Eds.), *Moral inferences.* Psychology Press.

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., & Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLoS ONE, 8*(4), e56515.

Sovacool, B. K. (2008). Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? *Journal of Bioethical Inquiry*, *5*, 271–282.

Stephan, P. E. (2012). *How economics shapes science*. Harvard University Press.

Strevens, M. (2011). Economic approaches to understanding scientific norms. *Episteme*, *8*(2), 184–200.

van Bavel et al. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour, 4*, 460–471.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspective on Psychological Science, 13*(4), 411–417.

Whyte, K. P., & Crease, R. P. (2010). Trust, expertise, and the philosophy of science. *Synthese*, *177*(3), 411–425.

World Health Organization. (2020, October 15). Behavioural considerations for acceptance and uptake of COVID-19 vaccines. Retrieved from https://www.who.int/publications/i/item/9789240016927

Wiggins, B. J., & Chrisopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202.

Wilholt, T. (2013). Epistemic trust in science. *The British Journal for the Philosophy of Science*, *64*(2), 233–253.

Williamson, T. (2002). *Knowledge and its limits*. Oxford University Press.

Ziliak, S., & McCloskey, D. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press.

Zollman, K. J. S. (2018). The credit economy and the economic rationality of science. *Journal of Philosophy, 115*(1), 5–33.