



an open access  journal



Citation: Hug, S. E. (2022). Towards theorizing peer review. *Quantitative Science Studies*, 3(3), 815–831.
https://doi.org/10.1162/qss_a_00195

DOI:
https://doi.org/10.1162/qss_a_00195

Peer Review:
https://publons.com/publon/10.1162/qss_a_00195

Received: 27 January 2022
Accepted: 10 April 2022

Corresponding Author:
Sven E. Hug
sven.hug@uzh.ch

Handling Editor:
Ludo Waltman

Copyright: © 2022 Sven E. Hug.
Published under a Creative Commons
Attribution 4.0 International (CC BY 4.0)
license.



RESEARCH ARTICLE

Towards theorizing peer review

Sven E. Hug 

Department of Psychology, Social and Business Psychology, University of Zurich,
Binzmühlestrasse 14/13, 8050 Zurich, Switzerland

Keywords: academic peer review, grant funding, peer review theory, scholarly communication, science of science

ABSTRACT

Academic peer review is seriously undertheorized because peer review studies focus on discovering and confirming phenomena, such as biases, and are much less concerned with explaining, predicting, or controlling phenomena on a theoretical basis. In this paper, I therefore advocate for more theorizing in research on peer review. I first describe the main characteristics of the peer review literature, which focuses mainly on journal and grant peer review. Based on these characteristics, I then argue why theory is useful in research on peer review, and I present some theoretical efforts on peer review. I conclude by encouraging peer review researchers to be more theoretically engaged and outline activities that theoretical work on peer review could involve. This invitation to theory-building complements recent roadmaps and calls that have emphasized that we need to have better access to peer review data, improve research design and statistical analysis in peer review studies, experiment with innovative approaches to peer review, and provide more funding for peer review research.

1. INTRODUCTION

Academic peer review in its modern form and purpose emerged in the late 1960s and early 1970s (Baldwin, 2018, 2020; Moxham & Fyfe, 2018). Since then, a large literature on peer review has been published (Batagelj, Ferligoj, & Squazzoni, 2017; Grimaldo, Marušić, & Squazzoni, 2018). Probably the most common finding in this literature is that scholars who review the research of their colleagues often arrive at very different judgments (e.g., Bornmann, 2011; Cicchetti, 1991; Lee, Sugimoto et al., 2013). Low interrater reliability could therefore be considered a hallmark of peer review. Despite the robustness and repeated replication of this disagreement effect, few studies have investigated the reasons behind it (Bornmann, Mutz, & Daniel, 2010; Seeber, Vlegels et al., 2021) and research is needed to identify the factors contributing to this phenomenon (Hesselberg, Fostervold et al., 2021). It seems therefore that research on peer review does focus more on discovering and confirming effects and less on explaining effects and building theories. In fact, several authors have concluded that there is “a stark discrepancy between the number of empirical peer review studies and the theoretical understanding of the process” (Gläser & Laudel, 2005, p. 187); that most studies fail to relate their empirical findings to theory (Bornmann, 2008); that there are only folk theories on peer review that explain little (Reinhart, 2017); that a comprehensive model of peer review and its broader context is essential but lacking (Chubin & Hackett, 1990); and that

there is a general lack of theory in research on peer review (Hirschauer, 2004)¹. Furthermore, Elson, Huff, and Utz (2020) called for field experiments to increase knowledge on mechanisms and determinants of peer review processes, Johnson and Hermanowicz (2017) concluded that explanatory mechanisms need to be incorporated more in analyses of peer review, and Reinhart and Schendzielorz (2021) underlined that we need a more comprehensive understanding of how peer review works. Despite this clear theoretical deficit of research on peer review, theorizing and theory are not desiderata in recent roadmaps and calls to action (Azoulay & Li, 2020; Bendiscioli, 2019; Bendiscioli, Firpo et al., 2021; Ioannidis, Berkwitz et al., 2019; Lauer & Nakamura, 2015; Lee & Moher, 2017; Rennie, 2016; Severin & Egger, 2021; Squazzoni, Ahrweiler et al., 2020; Tennant, Dugan et al., 2017; Tennant & Ross-Hellauer, 2020). In this paper, I therefore advocate for more theoretical engagement in research on peer review.

Because I invite researchers to be more theoretically engaged, I have to make clear what I understand by theorizing and theory. But defining these notions is not straightforward, which might have been the reason why the authors mentioned above pointing to a theoretical deficit did not state what they mean by theory. My understanding of the two notions is as follows. A *theory* is a set of linked propositions or concepts that explain, predict, or control one or several phenomena (Borsboom, van der Maas et al., 2021; Haslbeck, Ryan et al., 2021). Drawing on Woodward (1989), I understand phenomena as (relatively) stable and robust features of peer review. I am agnostic about the scope of a theory: It can be narrow (e.g., focusing on an aspect of a particular peer review procedure or a single phenomenon) or broad (e.g., focusing on a type of peer review or peer review in general). *Theorizing* is the process of creating new theories (generate, construct, build) and modifying existing theories (adapt, develop, improve). Naturally, this process can also generate theoretical contributions that do not have the status of a theory but serve as steps on the way to a theory (e.g., descriptions, definitions, categorizations, concepts, hypotheses, frameworks, taxonomies, models). Theorizing can be done in many ways (e.g., through induction, deduction, abduction), and it also involves testing and evaluating theories. I recognize that my understanding of theory is very specific and that there are many other definitions and uses of this notion. For example, Abend (2008) identified seven meanings of theory in sociology alone. My perspective on theory and theorizing is therefore not meant to exclude other views but to make clear from which standpoint I argue.

As scholars from many disciplines are involved in research on peer review and as peer review research is highly fragmented (see Section 2), I think that a more inclusive and integrative approach is essential to advance research on peer review. In this paper, I will therefore attempt to identify the main characteristics of the literature on peer review—or the lowest common denominators of peer review research—and, based on these characteristics, I will argue for more theoretical engagement. This means that I will avoid basic arguments (e.g., theory building is fundamental to the expansion of knowledge) or topical arguments (e.g., theorizing is one of the remedies for the replication crisis), and I will not base my rationale on a particular method (e.g., computational modeling) or specific topics and issues (e.g., innovations in peer review, disagreement effect). In line with my approach, I will therefore first describe the main characteristics of the literature (Section 2): The literature focuses on *phenomena* and the *meritocratic legitimacy* of peer review, and it is *application oriented* and

¹ Mitroff and Chubin (1979) were perhaps the first authors to criticize the low level of theoretical engagement when they discussed two studies on peer review at the National Science Foundation and concluded that one study was “essentially atheoretical” (p. 224), while the other did not include enough theoretical perspectives.

fragmented. Based on these characteristics, I will then argue why theory can be useful in research on peer review and describe further, less prominent characteristics of the literature (Section 3). While several authors have noted that there is a theory gap, it has not been pointed out that theoretical contributions do exist. One could thus conclude that the literature is completely devoid of theoretical efforts. To illustrate that this is not the case and to provide references for readers interested in peer review theory, I will present some publications that could be considered theoretical (Section 4). I will conclude by encouraging peer review researchers to be more theoretically engaged and outline activities that theoretical work on peer review could involve (Section 5). Note that when I use the term “peer review” in this paper, I always imply that peer review is a highly diverse (social) practice, is institutionalized in various ways, includes many different procedures, serves different purposes, and evolves through time. While this paper is based mainly on literature on journal and grant peer review, I recognize that there are many other uses of peer review in academia (e.g., in hiring and promotion processes, institutional evaluations, or awarding prizes).

2. MAIN CHARACTERISTICS OF THE PEER REVIEW LITERATURE

I suggest that the literature on peer review has four main characteristics. First, it focuses on phenomena. The theory gap identified by several authors (see Section 1) can be readily verified by consulting literature reviews, as they neither contain peer review theories nor identify other theoretical contributions (e.g., Bornmann, 2011; Campanario, 1998a, 1998b; Godlee & Jefferson, 2003; Guthrie, Ghiga, & Wooding, 2018b; Lee et al., 2013; Sabaj Meruane, González Vergara, & Pina-Stranger, 2016; Tennant & Ross-Hellauer, 2020; Weller, 2001). Through the lens of my understanding of theory, however, we can see that the literature is very much engaged with one element of theory, namely, the examination of phenomena (i.e., relatively stable and robust features of peer review). Probably the most prominent class of phenomena in research on peer review are biases (for an overview, see Lee et al. [2013]). While some of these phenomena are well established, uncontested, and supported by robust evidence, such as the disagreement effect², others are highly contested, such as gender bias (Sato, Gyax et al., 2021)³, or there is only preliminary and suggestive evidence—for example, on conservatism (Franzoni, Paula, & Veugelers, 2022; Guthrie et al., 2018b). In addition to biases, many other supposed, corroborated, or obvious phenomena can be found in the literature, such as overburdening (Kovanis, Porcher et al., 2016) or lack of transparency (Horbach, Hepkema, & Halfman, 2020; Ross-Hellauer, 2017). The focus on phenomena also becomes apparent when one considers that the development of theories for explaining phenomena, such as conservatism or gender bias, has only just commenced. For example, Gross and Bergstrom (2021) proposed a theory to explain why ex-post peer review encourages high-risk research, while ex-ante review discourages it, and the GRANteD project aims to determine causes and effects of gender bias in grant funding (<https://www.granted-project.eu>). Based on these considerations, I conclude that most of the literature focuses on discovering, confirming, and examining phenomena and is much less concerned with explaining, predicting, or controlling phenomena on a theoretical basis. This conclusion is the reason why I call peer

² Erosheva, Martinková, and Lee (2021) recently argued that measures of interrater reliability (IRR) calculated from range-restricted data, which is often the case in peer review studies, are not valid. Using two data sets that contained ratings across the complete range of grant proposals, they found that reviewer agreement is *good* (IRR values of 0.61 and 0.64, respectively). In contrast, Bornmann et al. (2010) reported a *low* level of IRR (0.34) for journal peer review in their meta-analysis. The study by Erosheva et al. (2021) therefore fundamentally challenges the evidence on the probably most robust phenomenon in peer review research.

³ See also the debate of Bornmann, Mutz, and Daniel (2007) and Marsh, Bornmann et al. (2009), as well as the debate of Squazzoni (2021) and Hagan (2021) on the findings of Squazzoni, Bravo et al. (2021).

review research undertheorized: There is an abundance of research on phenomena but a paucity of studies that explain, predict, or control these phenomena. As we will see in the next paragraph, phenomena in peer review research are mostly shaped by meritocratic principles. Note that peer review studies often seem to explain or predict. Take, for example, a study that predicts or explains publication decisions of submitted manuscripts on the basis of features of the authors (e.g., age, gender) or reviewer ratings (e.g., originality, rigor, relevance). Such a study does not correspond to my understanding of theory because the relationship of these variables only enables us to reveal phenomena (e.g., age bias, gender bias, conservatism) but does not explain or predict phenomena on a theoretical basis (Bogen & Woodward, 1988; Borsboom et al., 2021; Woodward, 1989). And such studies normally do not describe themselves as theoretical, either. In my understanding, a theory requires a set of linked propositions that explain, predict, or control phenomena—and a mere group of factors or variables is obviously not a system of propositions. Note also that peer review studies often intervene or recommend controlling, changing, or improving certain aspects of peer review but the (recommended) interventions are often not theoretically grounded. For example, none of the interventions included in the systematic reviews by Bruce, Chauvin et al. (2016) and Recio-Saucedo, Crane et al. (2022) seem to be theory based. As a consequence, the recommended interventions often remain too vague and unspecific to be useful in practice, or it is unclear in which settings and under which conditions they will work. For example, we recommended that funding agencies should train reviewers from the humanities to use the same criteria, as there seem to be two criteria norms among humanities scholars (Hug & Ochsner, 2022), but neither did we specify what exactly should be taught, how it should be taught, and what the precise benefits of such a training would be, nor did we connect the recommendation to the broader topic of reviewer training (Callaham, 2003; Chong, 2021; Hesselberg, Dalsbø et al., 2020).

Second, the literature focuses on the meritocratic legitimacy of peer review. The ideal of meritocracy is a guiding principle in academia (Merton, 1973; Scully, 2002; van den Brink & Benschop, 2012). Accordingly, evaluation and decision-making procedures must be consistent with meritocratic principles to be perceived as legitimate by researchers (Posselt, Hernandez et al., 2020). Thorngate, Dawes, and Foddy (2009) defined five standards for meritocratic assessments: Merit judgements must be efficient, consistent, equitable, valid, and transparent to be perceived as legitimate⁴. Four of these five standards, or meritocratic properties, correspond to features of the peer review literature. Specifically, there is a general consensus that the bulk of research “casts peer review as an instrument or test that has to be evaluated with respect to efficiency, reliability, fairness, and (predictive) validity” (Hug & Aeschbach, 2020, p. 13; see also Bedeian, 2004; Bornmann, 2011; Bornstein, 1991; Butchard, Rowberry et al., 2017; Daniel, 1993; Guthrie, Ghiga, & Wooding, 2018a; Hirschauer, 2004; Marsh, Jayasinghe, & Bond, 2008; Reinhart, 2012; Weller, 2001; Wood & Wessely, 2003)⁵. I hasten to add that the fifth standard of Thorngate et al. (2009), transparency, is becoming increasingly important in peer review research as open peer review is “a hot topic with a rapidly growing literature” (Ross-Hellauer, 2017, p. 3). I therefore conclude that research on peer review mainly examines whether peer review conforms to meritocratic principles and in this way assesses the (meritocratic) legitimacy of peer review. Let me reframe this using the legitimacy framework

⁴ Thorngate et al. (2009) use the term *fairness* instead of *legitimacy*. For definitions of legitimacy, see Johnson, Dowd, and Ridgeway (2006), Schoon (2022), and Tyler (2006).

⁵ This characterization of the literature is generally consistent with meritocratic ideals of evaluation prevalent in academia, such as the merit model (Mitroff & Chubin, 1979), the fairness doctrine (Peters & Ceci, 1982), or the ideal of impartiality (Lee et al., 2013).

of Schoon (2022). Peer review is the *object of legitimacy*, the scientific community is the *audience* that evaluates the object, and the *expectations* towards this object are efficiency, reliability, fairness, etc., which are considered desirable and appropriate by the audience (*assent*). Finally, the literature on peer review reflects that the audience examines whether the object *conforms* to these expectations. My considerations are generally in agreement with Lee et al. (2013) and Scully (2015). Lee et al. (2013) argued that impartiality legitimizes peer review outcomes, content, and institutions. Note that, in contrast to me, Lee and colleagues focus on one meritocratic principle (impartiality) and different types of legitimacy (psychological, social, and epistemic legitimacy). In her general observations on meritocracy, Scully (2015) noted that social systems are often assessed and discussed with respect to three principles of meritocracy, thereby “fine-tuning a meritocracy” (Scully, 2015, p. 1). I think this is in fact reflected in most of the peer review literature, but the literature focuses on two of the three principles, namely, whether peer review is an *appropriate measure of merit* (first principle) and whether *biases compromise equality of opportunity* (second principle). Although “fine-tuning” might imply small changes, I also understand more significant changes by this term, for example, those currently discussed and summarized as “peer review innovations” (Barroga, 2020; Bendiscioli & Garfinkel, 2021; Björk & Hedlund, 2015; Buckley Woods, Brumberg et al., 2022; Burley, 2017; Guthrie, 2019; Kaltenbrunner, Pinfield et al., 2022; Tennant et al., 2017). Based on the considerations in this paragraph, I update in which respect I consider peer review research undertheorized: There is an abundance of research on a certain type of phenomena (i.e., those related to meritocratic principles) but a paucity of studies that explain, predict, or control these phenomena.

Third, the literature is application oriented. I have already mentioned an important attribute above that indicates the applied nature of research on peer review (i.e., the focus on the meritocratic legitimacy), and I will add further attributes here. Reinhart and Schendzielorz (2021, p. 2) argued that research on peer review “started as a reaction to public criticism [...] in the 1970s” and “has retained a focus on perceived deficits and ways to improve on them up to the present.” This improvement focus can readily be verified by consulting literature from different decades. For example, Mahoney (1982, p. 220) mentioned that the authors of a study “seem to imply that we should be striving to increase objectivity and reliability in the peer-review process,” Cicchetti (1991, p. 119) made several suggestions “for improving the reliability and the quality of peer review,” and one of the aims of the edited volume by Godlee and Jefferson (2003) is to look at ways to improve peer review. I suggest that improvement is in fact so central in the literature nowadays that improvement has become an integral part of those three communicative moves identified by Swales (1990) that are used in the introduction of scholarly articles to establish the relevance of a study (i.e., establishing a territory, establishing a niche, occupying the niche). A move is “a segment of text that is shaped and constrained by a specific communicative function” (Holmes, 1997, p. 325). Specifically, I suggest that Swales’ three moves are often implemented as follows. The first move emphasizes the centrality or ubiquity of peer review, such as in allocating funding or journal space, in certifying knowledge claims, or for the scientific endeavor in general (establishing a territory). The second move recognizes one or several meritocratic properties of peer review to be suboptimal (establishing a niche), and the third move presents or suggests a solution for the suboptimal properties (occupying the niche). To illustrate these moves, I use an article reporting an experiment that compared the performance of panel peer review and distributed peer review in allocating telescope time at the European Southern Observatory (Kerzendorf, Patat et al., 2020). The article highlights the centrality of peer review (“peer review of proposals for the allocation of resources is a foundation of modern science”), directs the attention to suboptimal

properties caused by a very high number of applications (“the heavy load [on the panel] has severe consequences on the review quality and the feedback that is provided to the applicants”), and presents a solution (“[machine-learning enhanced] Distributed Peer Review promises to alleviate several of the described problems”) (all quotes from Kerzendorf et al., 2020, p. 711). I therefore conclude that the literature puts a strong emphasis on *intervening* in practice and *improving* peer review with respect to meritocratic properties. I consider this conclusion to be consistent with Scully’s (2015) notion of “fine-tuning a meritocracy” (see above). In addition to the improvement focus, the article by Kerzendorf and colleagues illustrates that many peer review studies are *local*; that is, they focus on a particular journal, funding instrument, discipline, or research community (here the allocation of telescope time at the European Southern Observatory). The article also illustrates that peer review studies are often conducted by *insiders*; that is, researchers examine peer review processes in their own field (here astronomers and astrophysicists) and publish the findings in a journal of this field (here *Nature Astronomy*). Note that I have derived the two characteristics *local* and *insider* from Weller (2001, p. 8) who concluded that research on (journal) peer review is scattered and “does exist in almost every scholarly field with a journal publication outlet” and from Hirschauer (2004), who argued that peer review studies are mainly conducted by researchers who have not been trained to observe their own research practice. The three characteristics of the literature discussed in this paragraph suggest that most of the research is carried out locally by insiders to improve peer review practice in their community⁶.

Fourth, the literature is fragmented. Grimaldo et al. (2018) conducted a bibliometric analysis of the peer review literature and found a fragmentation in terms of researchers (i.e., many small coauthorship clusters) and knowledge (i.e., many small cocitation clusters). Other authors previously pointed out that the literature is fragmented, but without providing evidence (e.g., Campanario, 1998a; Largent & Snodgrass, 2016; Weller, 2001). I therefore suggest that fragmentation is the fourth characteristic of the literature. While Grimaldo et al. (2018) attributed the fragmentation to limited access to peer review data and to the lack of funding of peer review research, my characterization of the literature suggests two additional factors that could have facilitated the fragmentation. On the one hand, many studies are interested in examining and fine-tuning meritocratic properties of local peer review processes only (i.e., of a particular journal or funding instrument, within a particular discipline or research community; see also note 6). On the other hand, theoretical engagement is low, which hinders the creation of a common vocabulary and a shared knowledge base.

3. WHY IS THEORY USEFUL?

Based on the main characteristics of the literature described above, I will first argue why theory can be useful in research on peer review. I will then describe further, less prominent characteristics of the literature and provide more reasons for the usefulness of theory. We have seen that the literature puts a strong emphasis on intervening in practice and improving peer review with respect to meritocratic properties. This presupposes that one knows or investigates *how* peer review works and *how* the meritocratic properties can be fine-tuned. However, we have

⁶ The local focus of most studies might also reflect the self-governance and autonomy of scholarly communities and disciplines; that is, scholars examine and fine-tune meritocratic properties of peer review in their own community to self-control and legitimize their evaluation practices. Declining legitimacy seems indeed to have been a motivation for the study by Kerzendorf et al. (2020, p. 1), as they pointed out that the issues with the existing procedure “contribute to increasing levels of frustration in the community and to the loss of credibility in the whole selection process.” It will be interesting to see whether local peer review studies will still be done in the future or whether they will be replaced by metascience studies (see note 7).

noted that the literature focuses on examining meritocratic phenomena—and is less concerned with the *how* (i.e., explaining, predicting, or controlling these phenomena). We should thus place greater emphasis on the *how* and theorize the mechanisms that generate the phenomena. This would benefit the applied goal of peer review research and facilitate interventions in practice. In addition to the applied nature of peer review research and its focus on phenomena, we have seen that the literature is fragmented. I think that a more theoretical approach can reduce the fragmentation of knowledge and researchers. Specifically, extant and new results from local studies could be integrated in theories, and in this way, it could be assessed to which extent the results generalize beyond the local context. Theories could also diminish the fragmentation of future research, as theories could be used to evaluate which research questions and studies are worthy of pursuit. Moreover, theories could provide a common vocabulary and a shared knowledge base and thus facilitate communication among researchers and foster their collaboration. In this way, theory-building could help to establish Peer Review Studies as a new and interdisciplinary research field, which, according to Squazzoni, Brezis, and Marušić (2017) and Tennant and Ross-Hellauer (2020), is desirable.

I have suggested above that most of the literature focuses on local peer review processes (e.g., of a particular journal or funding instrument, within a particular discipline) and studies are often conducted by insiders. Since the last decade, however, we have been able to observe a shift towards more *global, transdisciplinary* research that is conducted by *outsiders* (i.e., researchers examine peer review processes outside their own field). For example, the COST Action New Frontiers of Peer Review (PEERE) was launched “to improve efficiency, transparency, and accountability of peer review through a trans-disciplinary, cross-sectoral collaboration” (COST, 2013, p. 2), the name of the International Congress on Peer Review and Biomedical Publication was changed “to replace ‘biomedical’ with ‘scientific’ in an effort to broaden the scope and engage researchers, editors, and others in all sciences” (Rennie & Flanagan, 2018, p. 350), and the Research on Research Institute was founded in 2019 because research on how research is funded, practiced, and evaluated “is often poorly joined-up” (RoRI, 2021). The shift away from local peer review processes is also reflected in recent large-scale studies that analyzed facets of peer review across disciplines (e.g., Horbach et al., 2020; Squazzoni et al., 2021; van den Besselaar, Sandström, & Schiffbaenker, 2018)⁷. I think that this new brand of peer review research requires a more theoretical approach to build frameworks that can compare and contrast facets of peer review across contexts (disciplines, purposes, regions, etc.). In a cross-disciplinary analysis of review reports from 740 journals, Garcia-Costa, Squazzoni et al. (2022, p. 1) arrived at a similar conclusion: “[...] increasing the standards of peer review at journals requires effort to assess interventions and measure practices with context-specific and multidimensional frameworks.” A more theoretical approach would thus advance our understanding of the contexts and conditions in which improvements and innovations of peer review are successful.

I have thus far focused on what could be called the *meritocratic paradigm* of peer review research, but there is, of course, research on peer review beyond this paradigm. For example, Lamont (2009), Reinhart (2012), and Derrick (2018) examined, simply put, how peer review as a process resolves disagreement and produces evaluation outcomes perceived as legitimate.

⁷ Possible reasons for this shift include peer review data that are increasingly available in digital form and in large quantities as well as the emergence of the metascience movement that “produces quantitative studies meant to describe and evaluate science on a macro scale” to “motivate reforms in scientific practice” across disciplines (all quotes from Peterson & Panofsky, 2020, p. 3). In fact, the RoRI mentioned above as an example of the shift was one of the organizers of the Metascience 2021 Conference (<https://metascience2021.org/about/>).

Van den Brink and Benschop (2012) analyzed how the notion of academic excellence is generated in the evaluation of professorial candidates. Paltridge (2017) studied reviewers' reports through different linguistic lenses. Hamann and Beljean (2021) compared gatekeeping processes in academia and the stand-up comedy industry. And several scholars investigated the history of peer review (e.g., Baldwin, 2015; Biagioli, 2002; Burnham, 1990; Hooper, 2019; Moxham & Fyfe, 2018; Newman, 2019; Zuckerman & Merton, 1971). But research on peer review beyond the meritocratic paradigm is difficult to characterize because there is considerably less such research and it is also fragmented. In a laudable attempt to describe the approaches that scholars have used to analyze academic evaluation, including but not restricted to peer review, Hamann and Beljean (2017) identified five perspectives (functionalist, power-analytical, performative, social-constructivist, pragmatist). These perspectives, however, were not described in detail, and the authors themselves assessed them as tentative and far from being distinct or mutually exclusive. I am therefore not able to point out how exactly theorizing might be useful for research beyond the meritocratic paradigm other than what I have already mentioned: Theories could help to reduce the fragmentation of research. But perhaps arguing why theory is useful for research beyond the meritocratic paradigm is not an urgent matter, as some of this research seems to have a theoretical focus already. For example, Chubin and Hackett (1990) argued that a comprehensive *model* of peer review and its broader context is essential. Other authors also emphasized that the broader context of peer review—and the relationship between the context and peer review—needs to be considered and theorized (Mitroff & Chubin, 1979; Neidhardt, 2016; Reinhart, 2012; Reinhart & Schendzielorz, 2021). Here, *broader context* refers to systems in which peer review is embedded, such as organizations, government, science, society, economy, or culture. While I have called research within the meritocratic paradigm undertheorized because there are many studies on phenomena but few that explain, predict, or control these phenomena, I consider many *aspects* of peer review (e.g., those related to the broader context) to be undertheorized simply because there is little research beyond the meritocratic paradigm that could have addressed these aspects.

While I keep emphasizing that research on peer review is undertheorized, one may object that theorizing is not necessary because theories already exist—at least implicitly or tacitly. This objection is consistent with Mitroff and Chubin (1979, p. 219), who pointed out in the early days of peer review research that data and methods are inseparable from theory: “[...] data can neither be collected in the first place, nor analyzed in the second, apart from some prior theoretical point of view. That is, one does not collect data without having presupposed some hypothesis, theory, or model, no matter how implicit, unconscious, or informal it may be.” From my perspective, tacit theories do not render theorizing obsolete but change its focus: Theorizing would thus become the process of making tacit assumptions and models explicit. In this way, other researchers previously unaware of tacit theories would be enabled to appraise them and build on them.

I have based my arguments for more theoretical engagement on the main characteristics of the peer review literature because I consider a holistic, inclusive, and integrative perspective essential to reduce the fragmentation of peer review research. My arguments are thus rather general. However, we can make more palpable arguments for theorizing if we focus on more concrete issues in peer review research. Although this is beyond the scope of the paper, I will provide one such argument. Take, for example, the complexity of peer review, which was linked to theory by Mitroff and Chubin already more than 40 years ago (Mitroff & Chubin, 1979, p. 224: “something so complex as peer review requires simultaneous and explicit examination from a number of diverse and competing theoretical perspectives”). Using recent research, we can illustrate why theorizing and theory are indispensable for addressing the

complexity of peer review. Recio-Saucedo et al. (2022) conducted a realist synthesis of 50 interventions in grant peer review and concluded that “changes that worked for a funder created new or exacerbated existing issues for other stakeholders [e.g., higher education institutions, applicants]” (p. 24). Kaltenbrunner et al. (2022) provided an analytical overview of innovations in journal peer review and found that “peer review innovations partly pull in mutually opposed directions” (p. 1). Hence, if we do not start to theorize the respective phenomena and how they are interrelated, interventions in peer review practice will remain random and futile.

4. SOME THEORETICAL EFFORTS ON PEER REVIEW

Several authors have noted that there is a theory gap (see Section 1), and I have also underlined that peer review is undertheorized. But so far, nobody has pointed out that theoretical contributions do exist. One could thus conclude that the literature is completely devoid of theoretical efforts, which is, of course, not the case. The purpose of this section is therefore to show that there are contributions that could be considered theoretical, but the purpose is not to provide an exhaustive overview⁸. This means that the studies included in this section are not the result of a systematic search and appraisal but represent an eclectic collection that focuses on recent studies⁹, includes research from a broad range of disciplines, and overrepresents research beyond the meritocratic paradigm. Moreover, some of the publications I have already cited in this paper could be considered theoretical as well, but for the sake of diversity, I have included other studies here. I will subsume the studies under six questions: Why and how has peer review evolved? What are researchers’ expectations and perceptions of peer review? How can single peer review phenomena be explained? What is the evidence for peer review phenomena and for the relationship between components of peer review and phenomena? Why and how does peer review work? How is peer review related to its contexts? I will outline why I consider the studies subsumed under each question to be theoretical, but I will not describe the theoretical characteristics of each study in detail. Note that I consider the studies summarized below to be in line with my understanding of theory or theorizing.

4.1. Why and How Has Peer Review Evolved?

None of the three studies included here has an overt theoretical ambition or presents a theory, but I read them as attempts to explain why and how journal peer review has evolved. In addition, the first study contains a prediction about the future evolution of peer review. Pontille and Torny (2015) studied the diversification of judging instances in journal peer review since the 17th century (editor-in-chief, editorial committee, external referee) and explained the evolution and diversity of peer review processes as the result of the growing number of publications and the two often conflicting needs for fast dissemination and validation of knowledge. They argued that today’s configuration of dissemination and validation has enabled readers to become a new key judging instance, which has the potential to transform the whole review process. Baldwin (2018) showed that it was only in the late 20th century that peer review came to be seen as a process central to scientific practice and that this perception can be traced to hearings in the United States in 1975 in which various stakeholders sought to navigate a growing tension between desires for scientific autonomy and public accountability in

⁸ For example, I have included just one study from philosophy of science (Arvan, Bright, & Heesen, 2022) and left out others (e.g., Avin, 2019; Heesen, 2018; Lee, 2012).

⁹ The first theoretical contribution to research on peer review was perhaps made in the late 1960s. Stinchcombe and Ofshe (1969) proposed a formal model of the editorial review process and predicted that nearly half of the good papers submitted to a journal will be rejected.

controversies over government science funding. Merriman (2021) analyzed when major elements of peer review emerged in journals of the American Sociological Association and argued that the ongoing evolution of peer review in these journals has not been driven by epistemic considerations but rather by “efforts to steward the scarce attention of editors while preserving an open submission policy that favors the authors’ interests” (Merriman, 2021, p. 341).

4.2. What Are Researchers’ Expectations and Perceptions of Peer Review?

While the following four studies do not frame themselves as theoretical, they uncover and conceptualize the researchers’ intuitive understanding and expectation of peer review, which I consider a theoretical contribution for two reasons. First, concepts are part of my definition of theory and, second, I see these conceptualizations as a first step in designing practical interventions, for example, to reduce reviewer burden or to motivate researchers to review. Tercier and Callaham (2007) interviewed 72 referees from medicine on their beliefs about the review process to generate a normative model of journal peer review. The model comprises four domains (manuscript, review, reviewer, review process), each specified by desirable attributes. Glonti, Cauchi et al. (2019) conducted a scoping review on the roles and tasks of referees in journal peer review in biomedicine. They compiled a list of 76 role-related statements and organized them into 13 themes, based on which they defined the requirements for an ideal peer reviewer. Severin and Chataway (2021) organized focus groups and found that scholars, referees, editors, and publishers believed overburdening to be caused by an increase in manuscript submissions, insufficient editorial triage, lack of reviewing instructions, difficulties in recruiting reviewers, inefficient manuscript handling, and a lack of institutionalization of peer review. In a scoping review, Mahmić-Kaknjo, Utrobičić, and Marušić (2021) identified 25 reasons and motivations for serving as a journal referee and organized the motivations into four categories using two dimensions (internal vs. external; incentives vs. disincentives).

4.3. How Can Single Peer Review Phenomena Be Explained?

The studies included here do have either a clear causal and explanatory focus or model aspects of peer review. Horbach and Halfman (2019) used a taxonomy of peer review procedures to assess the effectiveness of review models at 361 journals to detect erroneous or fraudulent research and found that author blinding, involving the wider community, using digital tools, constraining interaction between authors and reviewers, and conducting prepublication reviews are significantly more effective in preventing retractions than other elements of peer review. In a scoping review, Feliciani, Luo et al. (2019) identified 46 studies that represented elements of peer review in computational models (e.g., agent-based models, latent Markov models). Computational models are therefore likely the most common approach so far to theorize peer review. Instead of summarizing the phenomena and mechanisms that computational models focus on here, I refer the reader to the overviews by Feliciani et al. (2019) and Shah (2022). Arvan et al. (2022) introduced three assumptions that prepublication review is based upon (competency, intersubjectivity, atomism) and argued, based on these assumptions and modified Condorcet jury theorems, that a crowd-sourced model of postpublication review is likely to do better at sorting papers by quality than journal-solicited prepublication review. Roumbanis (2022) theorized how panels struggle to reach a consensus decision when there is strong disagreement among panelists. He calls the results of such struggles *agonistic chance*, that is, unforeseen consequences of social interactions in peer review. To explain agonistic chance, Roumbanis proposed and empirically tested a framework consisting of five concepts (evaluative crossroads, aporetic position, radical compromise, collective risk-

taking, fateful events). Gross and Bergstrom (2021) proposed a verbal theory and a corresponding formal model to understand how peer review shapes the questions researchers choose to study. The theory predicts that ex-post evaluation encourages high-risk research while ex-ante evaluation discourages it because “investigators can leverage the differences between their private beliefs and those of the community when peer reviewers evaluate a completed experiment, but they have no opportunity to leverage these differences when peers evaluate a proposed experiment” (Gross & Bergstrom, 2021, p. 6)¹⁰. Franzoni et al. (2022) devised a framework to analyze why funding agencies may eschew risky research. The framework includes factors within the research system that might contribute to risk aversion (accountability, short-term thinking, no tolerance for failure, bibliometric indicators, soft-money positions) and nine hypotheses on the behavior of principal investigators (refraining from submitting risky proposals, loss aversion), panelists (insurance agent view, bibliometric screening, risk-biased panelists), and funding agencies (no portfolio approach, interdisciplinary bias, review protocols concealing uncertainty, stress on agreement). The aim of the GRANted project is to clarify the concept of gender bias in grant funding, to provide empirical evidence on the prevalence of the phenomenon, and to determine causes and effects of gender bias (<https://www.granted-project.eu>). The project uses a heuristic model that focuses on the level of grant panels and includes mechanisms and processes related to the application, the selection of grants, and the effect of grants on careers (for details, see van den Besselaar, Mom et al., 2020). The RoRI CRITERIA project studies how review criteria influence gender inequalities in research funding (<https://researchonresearch.org/projects>) and models the causal relationships between features of the applicants (publication record, distinguished positions, institutional affiliation, academic rank and age, research field, gender), the funding agencies (eligibility criteria, review criteria, funding decision), and the quality of the proposals (Traag, 2021).

4.4. What Is the Evidence for Peer Review Phenomena and for the Relationship Between Components of Peer Review and Phenomena?

Guthrie et al. (2018a, 2018b) summarized and assessed the evidence from 105 publications on meritocratic properties of grant peer review (reliability, fairness, accountability, timeliness, etc.), and they categorized and organized the meritocratic properties on three levels of abstraction. While the review by Guthrie and colleagues is not theoretical in the sense that it focuses on explaining, predicting, or controlling, it assesses the robustness and stability of phenomena, which I understand to be a necessary component of a theory. As noted above, the evidence base for phenomena within the meritocratic paradigm is not always clear or strong, and it is therefore important to collect, categorize, and assess evidence on phenomena. Shepherd, Frampton et al. (2018) conducted an evidence synthesis of eight studies that evaluated the effects of innovations in grant peer review on various measures of effectiveness and efficiency. Although the authors have not used or created a theoretical framework for their synthesis, the study can be seen as a step towards a systematic integration of empirical findings on the (causal) relationship between types and components of peer review and meritocratic properties (fairness, reliability, efficiency, etc.) and might therefore be considered theoretical.

¹⁰ Evidence from grant peer review (e.g., Ayoubi, Pezzoni, & Visentin, 2021; Boudreau, Guinan et al., 2016) and journal peer review (Teplitskiy, Peng et al., 2021) seem to support the theory. The theory of Gross and Bergstrom (2021) was published on arXiv and submitted to *PNAS* before Teplitskiy et al. (2021) published their study on SSRN. The theory thus predicted a pattern that had not been observed.

4.5. Why and How Does Peer Review Work?

The following three studies have a clear theoretical ambition or framing and attempt to explain why and how peer review works. Mitroff and Chubin (1979) suggested that there are at least three models that could explain how referees evaluate and recommend funding applications: the accumulative advantage model, the political model, and the merit model (for details, see Mitroff & Chubin, 1979, pp. 219–225). Hirschauer (2019) theorized journal peer review as a content conflict (*Sachkonflikt*) that is guided by 12 tactics to intensify communication about a manuscript and to prevent the content conflict from escalating into a relationship conflict (*Beziehungskonflikt*) or a power conflict (*Machtkonflikt*). Hirschauer organized the tactics into four strategies: intensification of attention, objectivity, disagreement, and authorship. According to Hirschauer, strategies and tactics facilitate the improvement of a manuscript and configure peer review as a site where the opinions and positions of professionals on a manuscript are explicated (he calls this the performative publicity of peer review). Reinhart and Schendzielorz (2021) published what they call preliminaries to theories of peer review. They argued that peer review fulfills three interrelated roles: It is a mechanism to assess quality (process), to decide on scarce resources (outcome), and to self-govern science (context). Their considerations focus on process and context because most peer review studies address outcomes; that is, the validity, reliability, and fairness of decisions. Specifically, Reinhart and Schendzielorz proposed eight activities to describe peer review processes (postulation, consultation, decision, administration, discussion, presentation, observation, moderation), explained how peer review ensures the quality and legitimacy of judgments, and discussed peer review as a mode of (self-) governance of science.

4.6. How Is Peer Review Related to Its Contexts?

I am only aware of two studies that address the relationship between peer review and its broader context: the work by Reinhart and Schendzielorz (2021) discussed above and, tangentially, the framework of Langfeldt, Nedeva et al. (2020). Specifically, Langfeldt and colleagues proposed a framework to study context-specific understandings of research quality that consists of three dimensions: *types* and *attributes* of research quality notions, and organizational *sites* where notions are constituted, contested, and institutionalized. While the framework does not focus on peer review, it contextualizes peer review and provides a lens to analyze how notions of research quality used in peer review relate to quality notions in neighboring domains.

5. THEORIZE!

I conclude that despite more than 50 years of research on peer review and some theoretical efforts, peer review and its contexts remain seriously undertheorized. I therefore encourage peer review researchers to be more theoretically engaged. Based on my understanding of theory and theorizing, this would mean to work towards a set or sets of linked propositions that explain, predict, or control one or several peer review phenomena. However, because scholars from virtually all disciplines study peer review processes and their theoretical standpoints likely differ from mine, it seems sensible to me to be open, tolerant, and inclusive in terms of what theorizing means and how it is performed as well as in terms of what counts as theory and theoretical contribution. I therefore invite other researchers to add their theoretical standpoint—and to use their standpoint to theorize peer review. My call to theorize, however, is not a theory imperative. Not every contribution, not every paper, has to be theoretical,

linked to theory, or guided by theory—but overall, more theoretical engagement is certainly desirable and worthwhile to expand our knowledge on peer review.

From my perspective, theoretical work on peer review could involve at least six broad activities. The first activity, *defining academic peer review*, could involve reviewing, conceptualizing, and comparing what we mean by peer review in different scholarly contexts and in academia in general. The second activity, *identifying phenomena*, could entail discovering phenomena as well as describing, defining, and conceptualizing them clearly and precisely. It could also entail collecting evidence on phenomena, although this is rather an empirical than a theoretical activity. However, as the evidence base for peer review phenomena is not always clear or strong, it is important to establish the robustness and stability of phenomena. The third activity, *theory building and modification*, could involve generating, developing, and connecting concepts from scratch to explain phenomena. It could also involve applying and adapting theories from other domains to peer review, modifying and refining existing peer review theories, or making tacit assumptions and models explicit. Ideally, we value a definition, a hypothesis, or a minitheory about an aspect of peer review as much as a comprehensive taxonomy, a complex model, or a theory with a broad scope. And we should welcome verbal as well as formal theories. Naturally, all theories need to be empirically tested. As there are no clearly articulated research programs on peer review, a fourth activity could be to *define the main (theoretical) research questions* that we want to address either within or beyond the meritocratic paradigm¹¹. Another activity could be to systematically *review and synthesize all theoretical efforts* on peer review, as the literature is fragmented and nobody has summarized the literature with respect to theoretical contributions yet. Finally, we could *reflect* on how we theorize and what we count as theory and theoretical contribution to further our theoretical capabilities.

This invitation to be more theoretically engaged complements recent roadmaps and calls that have emphasized that we need to have better access to peer review data, improve research design and statistical analysis in peer review studies, experiment with innovative approaches to peer review, and provide more funding for peer review research (Azoulay & Li, 2020; Bendiscioli, 2019; Bendiscioli et al., 2021; Ioannidis et al., 2019; Lauer & Nakamura, 2015; Lee & Moher, 2017; Rennie, 2016; Severin & Egger, 2021; Squazzoni et al., 2020; Tennant et al., 2017; Tennant & Ross-Hellauer, 2020). It would be reasonable to follow these proposals if we want to advance our understanding of how peer review works, how it is related to its contexts, and how we can develop it further.

ACKNOWLEDGMENTS

I thank Martin Reinhart for inspiring me to think about theory in research on peer review, and I am immensely grateful to Flaminio Squazzoni, Julian Hamann, Vincent Traag, Klaus Jonas, and Martin Reinhart for their thoughtful comments and suggestions on earlier versions of this paper. Finally, I thank the reviewers for thoroughly engaging with the manuscript and for providing critical comments.

COMPETING INTERESTS

The author has no competing interests.

¹¹ For research on journal peer review, Tennant and Ross-Hellauer (2020) recently developed a detailed list of research questions and topics to be studied.

FUNDING INFORMATION

No funding was provided for conceiving and writing the paper.

DATA AVAILABILITY

No data have been used in this paper.

REFERENCES

- Abend, G. (2008). The meaning of “theory.” *Sociological Theory*, 26(2), 173–199. <https://doi.org/10.1111/j.1467-9558.2008.00324.x>
- Arvan, M., Bright, L. K., & Heesen, R. (2022). Jury theorems for peer review. *British Journal for the Philosophy of Science*. <https://doi.org/10.1086/719117>
- Avin, S. (2019). Centralized funding and epistemic exploration. *British Journal for the Philosophy of Science*, 70(3), 629–656. <https://doi.org/10.1093/bjps/axx059>
- Ayoubi, C., Pezzoni, M., & Visentin, F. (2021). Does it pay to do novel science? The selectivity patterns in science funding. *Science and Public Policy*, 48(5), 635–648. <https://doi.org/10.1093/scipol/scab031>
- Azoulay, P., & Li, D. (2020). Scientific grant funding. *National Bureau of Economic Research Working Paper Series, No. 26889*. <https://doi.org/10.3386/w26889>
- Baldwin, M. (2015). Credibility, peer review, and *Nature*, 1945–1990. *Notes and Records: The Royal Society Journal of the History of Science*, 69(3), 337–352. <https://doi.org/10.1098/rsnr.2015.0029>, PubMed: 26495581
- Baldwin, M. (2018). Scientific autonomy, public accountability, and the rise of “peer review” in the Cold War United States. *Isis*, 109(3), 538–558. <https://doi.org/10.1086/700070>
- Baldwin, M. (2020). Peer review. *Encyclopedia of the History of Science*. <https://doi.org/10.34758/srde-jw27>
- Barroga, E. (2020). Innovative strategies for peer review. *Journal of Korean Medical Science*, 35(20), e138. <https://doi.org/10.3346/jkms.2020.35.e138>, PubMed: 32449322
- Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: A network analysis of research on peer review. *Scientometrics*, 113(1), 503–532. <https://doi.org/10.1007/s11192-017-2522-8>, PubMed: 29056788
- Bedeian, A. G. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning & Education*, 3(2), 198–216. <https://doi.org/10.5465/amle.2004.13500489>
- Bendisoli, S. (2019). The troubles with peer review for allocating research funding. *EMBO Reports*, 20(12), e49472. <https://doi.org/10.15252/embr.201949472>, PubMed: 31680417
- Bendisoli, S., Firpo, T., Bravo-Biosca, A., Czibor, E., Garfinkel, M., ... Woods, H. B. (2021). The experimental research funder’s handbook. *RoRI Working Paper*. <https://doi.org/10.6084/m9.figshare.17102426.v1>
- Bendisoli, S., & Garfinkel, M. (2021). *Dealing with the limits of peer review with innovative approaches to allocating research funding*. EMBO.
- Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures*, 12(1), 11–45. <https://doi.org/10.1080/1045722022000003435>
- Björk, B.-C., & Hedlund, T. (2015). Emerging new methods of peer review in scholarly journals. *Learned Publishing*, 28(2), 85–91. <https://doi.org/10.1087/20150202>
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, 97(3), 303–352. <https://doi.org/10.2307/2185445>
- Bornmann, L. (2008). Scientific peer review: An analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture: Journal of the Sociology of Self-Knowledge*, VI, 23–38.
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197–245. <https://doi.org/10.1002/aris.2011.1440450112>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3), 226–238. <https://doi.org/10.1016/j.joi.2007.03.001>
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multi-level meta-analysis of inter-rater reliability and its determinants. *PLOS ONE*, 5(12), e14331. <https://doi.org/10.1371/journal.pone.0014331>, PubMed: 21179459
- Bornstein, R. F. (1991). The predictive validity of peer review: A neglected issue. *Behavioral and Brain Sciences*, 14(1), 138–139. <https://doi.org/10.1017/S0140525X00065717>
- Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 16(4), 756–766. <https://doi.org/10.1177/1745691620969647>, PubMed: 33593167
- Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science*, 62(10), 2765–2783. <https://doi.org/10.1287/mnsc.2015.2285>, PubMed: 27746512
- Bruce, R., Chauvin, A., Trinquart, L., Ravaut, P., & Boutron, I. (2016). Impact of interventions to improve the quality of peer review of biomedical journals: A systematic review and meta-analysis. *BMC Medicine*, 14(1), 85. <https://doi.org/10.1186/s12916-016-0631-5>, PubMed: 27287500
- Buckley Woods, H., Brumberg, J., Kaltenbrunner, W., Pinfield, S., & Waltman, L. (2022). Innovations in peer review in scholarly publishing: A meta-summary. *SocArXiv*. <https://doi.org/10.31235/osf.io/qaksd>
- Burley, R. (2017). Peer review in the 21st century. *Information Services & Use*, 37(3), 259–261. <https://doi.org/10.3233/ISU-170850>
- Burnham, J. C. (1990). The evolution of editorial peer review. *JAMA*, 263(10), 1323–1329. <https://doi.org/10.1001/jama.1990.03440100023003>, PubMed: 2406470
- Butchard, D., Rowberry, S., Squires, C., & Tasker, G. (2017). Peer review in practice. In S. Rayner & R. Lyons (Eds.), *Academic Book of the Future: BOOC*. UCL Press. <https://doi.org/10.14324/111.9781911307679.15>
- Callaham, M. L. (2003). The evaluation and training of peer reviewers. In F. Godlee & T. Jefferson (Eds.), *Peer Review in Health Sciences* (2nd ed., pp. 164–182). BMJ Books.

- Campanario, J. M. (1998a). Peer review for journals as it stands today—Part 1. *Science Communication*, 19(3), 181–211. <https://doi.org/10.1177/1075547098019003002>
- Campanario, J. M. (1998b). Peer review for journals as it stands today—Part 2. *Science Communication*, 19(4), 277–306. <https://doi.org/10.1177/1075547098019004002>
- Chong, S. W. (2021). Improving peer-review by developing reviewers' feedback literacy. *Learned Publishing*, 34(3), 461–467. <https://doi.org/10.1002/leap.1378>
- Chubin, D. E., & Hackett, E. J. (1990). *Peerless science. Peer review and U.S. science policy*. State University of New York Press.
- Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119–186. <https://doi.org/10.1017/S0140525X00065675>
- COST. (2013). *Memorandum of understanding for the implementation of a European Concerted Research Action Designated as COST Action TD1306: New Frontiers of Peer Review (PEERE)*. <https://www.cost.eu/actions/TD1306>
- Daniel, H. D. (1993). *Guardians of science: Fairness and reliability of peer review*. VCH Verlagsgesellschaft. <https://doi.org/10.1002/3527602208>
- Derrick, G. (2018). *The evaluators eye. Impact assessment and academic peer review*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-63627-6>
- Elson, M., Huff, M., & Utz, S. (2020). Metascience on peer review: Testing the effects of a study's originality and statistical significance in a field experiment. *Advances in Methods and Practices in Psychological Science*, 3(1), 53–65. <https://doi.org/10.1177/2515245919895419>
- Erosheva, E. A., Martinková, P., & Lee, C. J. (2021). When zero may not be zero: A cautionary note on the use of inter-rater reliability in evaluating grant peer review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184, 904–919. <https://doi.org/10.1111/rssa.12681>
- Feliciani, T., Luo, J., Ma, L., Lucas, P., Squazzoni, F., Marušić, A., & Shankar, K. (2019). A scoping review of simulation models of peer review. *Scientometrics*, 121(1), 555–594. <https://doi.org/10.1007/s11192-019-03205-w>, PubMed: 31564758
- Franzoni, C., Paula, S., & Veugelers, R. (2022). Funding risky research. *Entrepreneurship and Innovation Policy and the Economy*, 1, 103–133. <https://doi.org/10.1086/719252>
- García-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2022). Measuring the developmental function of peer review: A multi-dimensional, cross-disciplinary analysis of review reports from 740 academic journals. *PeerJ*, 10, e13539. <https://doi.org/10.7717/peerj.13539>, PubMed: 35694383
- Gläser, J., & Laudel, G. (2005). Advantages and dangers of “remote” peer evaluation. *Research Evaluation*, 14(3), 186–198. <https://doi.org/10.3152/147154405781776085>
- Glonti, K., Cauchi, D., Cobo, E., Boutron, I., Moher, D., & Hren, D. (2019). A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. *BMC Medicine*, 17(1), 118. <https://doi.org/10.1186/s12916-019-1347-0>, PubMed: 31217033
- Godlee, F., & Jefferson, T. (Eds.). (2003). *Peer review in the health sciences* (2nd ed.). BMJ Books.
- Grimaldo, F., Marušić, A., & Squazzoni, F. (2018). Fragments of peer review: A quantitative analysis of the literature (1969–2015). *PLOS ONE*, 13(2), e0193148. <https://doi.org/10.1371/journal.pone.0193148>, PubMed: 29466467
- Gross, K., & Bergstrom, C. T. (2021). Why ex post peer review encourages high-risk research while ex ante review discourages it. *Proceedings of the National Academy of Sciences*, 118(51), e2111615118. <https://doi.org/10.1073/pnas.2111615118>, PubMed: 34921115
- Guthrie, S. (2019). *Innovating in the research funding process: Peer review alternatives and adaptations*. AcademyHealth.
- Guthrie, S., Ghiga, I., & Wooding, S. (2018a). *What do we know about grant peer review in the health sciences? An updated review of the literature and six case studies*. RAND Corporation. <https://doi.org/10.7249/RR1822>
- Guthrie, S., Ghiga, I., & Wooding, S. (2018b). What do we know about grant peer review in the health sciences? [Version 2; referees: 2 approved]. *F1000Research*, 6(1335), 1–23. <https://doi.org/10.12688/f1000research.11917.2>, PubMed: 29707193
- Hagan, A. (2021). *Opinion: Peer review study compromises response to gender bias*. <https://www.the-scientist.com/news-opinion/opinion-peer-review-study-compromises-response-to-gender-bias-68401>
- Hamann, J., & Beljean, S. (2017). Academic evaluation in higher education. In J. C. Shin & P. Teixeira (Eds.), *Encyclopedia of international higher education systems and institutions* (pp. 1–7). Springer Netherlands. https://doi.org/10.1007/978-94-017-9553-1_295-1
- Hamann, J., & Beljean, S. (2021). Career gatekeeping in cultural fields. *American Journal of Cultural Sociology*, 9(1), 43–69. <https://doi.org/10.1057/s41290-019-00078-7>
- Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2021). Modeling psychopathology: From data models to formal theories. *Psychological Methods*. <https://doi.org/10.1037/met0000303>, PubMed: 34735175
- Heesen, R. (2018). When journal editors play favorites. *Philosophical Studies*, 175(4), 831–858. <https://doi.org/10.1007/s11098-017-0895-4>
- Hesselberg, J. O., Dalsbø, T. K., Stromme, H., Svege, I., & Fretheim, A. (2020). Reviewer training for improving grant and journal peer review. *Cochrane Database of Systematic Reviews*, 11. <https://doi.org/10.1002/14651858.MR000056>
- Hesselberg, J. O., Fostervold, K. I., Ulleberg, P., & Svege, I. (2021). Individual versus general structured feedback to improve agreement in grant peer review: A randomized controlled trial. *Research Integrity and Peer Review*, 6(1), 12. <https://doi.org/10.1186/s41073-021-00115-5>, PubMed: 34593049
- Hirschauer, S. (2004). Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation. *Zeitschrift Fur Soziologie*, 33(1), 62–83. <https://doi.org/10.1515/zfsoz-2004-0104>
- Hirschauer, S. (2019). Urteilen unter Beobachtung. Performative Publizität im Peer Review. In S. Nicolae, M. Endress, O. Berli, & D. Bischur (Eds.), *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit* (pp. 275–298). Springer VS. https://doi.org/10.1007/978-3-658-21763-1_12
- Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16(4), 321–337. [https://doi.org/10.1016/S0889-4906\(96\)00038-5](https://doi.org/10.1016/S0889-4906(96)00038-5)
- Hooper, M. (2019). Scholarly review, old and new. *Journal of Scholarly Publishing*, 51(1), 63–75. <https://doi.org/10.3138/jsp.51.1.04>
- Horbach, S. P. J. M., & Halfman, W. (2019). The ability of different peer review procedures to flag problematic publications. *Scientometrics*, 118(1), 339–373. <https://doi.org/10.1007/s11192-018-2969-2>, PubMed: 30930504
- Horbach, S. P. J. M., Hepkema, W. M., & Halfman, W. (2020). The platform for responsible editorial policies: An initiative to foster

- editorial transparency in scholarly publishing. *Learned Publishing*, 33(3), 340–344. <https://doi.org/10.1002/leap.1312>
- Hug, S. E., & Aeschbach, M. (2020). Criteria for assessing grant applications: A systematic review. *Palgrave Communications*, 6, 37. <https://doi.org/10.1057/s41599-020-0412-9>
- Hug, S. E., & Ochsner, M. (2022). Do peers share the same criteria for assessing grant applications? *Research Evaluation*, 31(1), 104–117. <https://doi.org/10.1093/reseval/rvab034>
- Ioannidis, J. P. A., Berkwits, M., Flanagin, A., Godlee, F., & Bloom, T. (2019). The Ninth International Congress on Peer Review and Scientific Publication: A call for research. *JAMA*, 322(17), 1658–1660. <https://doi.org/10.1001/jama.2019.15516>, PubMed: 31524942
- Johnson, C., Dowd, T. J., & Ridgeway, C. L. (2006). Legitimacy as a social process. *Annual Review of Sociology*, 32(1), 53–78. <https://doi.org/10.1146/annurev.soc.32.061604.123101>
- Johnson, D. R., & Hermanowicz, J. C. (2017). Peer review: From “sacred ideals” to “profane realities.” In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 32, pp. 485–527). Springer. https://doi.org/10.1007/978-3-319-48983-4_10
- Kaltenbrunner, W., Pinfield, S., Waltman, L., Buckley Woods, H., & Brumberg, J. (2022). Innovating peer review, reconfiguring scholarly communication: An analytical overview of ongoing peer review innovation activities. *SocArXiv*. <https://doi.org/10.31235/osf.io/8hdxu>
- Kerzendorf, W. E., Patat, F., Bordelon, D., van de Ven, G., & Pritchard, T. A. (2020). Distributed peer review enhanced with natural language processing and machine learning. *Nature Astronomy*, 4, 711–717. <https://doi.org/10.1038/s41550-020-1038-y>
- Kovanis, M., Porcher, R., Ravaut, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. *PLOS ONE*, 11(11), e0166387. <https://doi.org/10.1371/journal.pone.0166387>, PubMed: 27832157
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press. <https://doi.org/10.4159/9780674054158>
- Langfeldt, L., Nedeve, M., Sörlin, S., & Thomas, D. A. (2020). Co-existing notions of research quality: A framework to study context-specific understandings of good research. *Minerva*, 58(1), 115–137. <https://doi.org/10.1007/s11024-019-09385-2>
- Largent, E. A., & Snodgrass, R. T. (2016). Blind peer review by academic journals. In A. S. Kesselheim & C. T. Robertson (Eds.), *Blinding as a solution to bias* (pp. 75–95). Academic Press. <https://doi.org/10.1016/B978-0-12-802460-7.00005-X>
- Lauer, M. S., & Nakamura, R. (2015). Reviewing peer review at the NIH. *New England Journal of Medicine*, 373(20), 1893–1895. <https://doi.org/10.1056/NEJMp1507427>, PubMed: 26559568
- Lee, C. J. (2012). A Kuhnian critique of psychometric research on peer review. *Philosophy of Science*, 79(5), 859–870. <https://doi.org/10.1086/667841>
- Lee, C. J., & Moher, D. (2017). Promote scientific integrity via journal peer review data. *Science*, 357(6348), 256–257. <https://doi.org/10.1126/science.aan4141>, PubMed: 28729501
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17. <https://doi.org/10.1002/asi.22784>
- Mahmić-Kaknjó, M., Utrobičić, A., & Marušić, A. (2021). Motivations for performing scholarly prepublication peer review: A scoping review. *Accountability in Research*, 28(5), 297–329. <https://doi.org/10.1080/08989621.2020.1822170>, PubMed: 32907396
- Mahoney, M. J. (1982). Publication, politics, and scientific progress. *Behavioral and Brain Sciences*, 5(2), 220–221. <https://doi.org/10.1017/S0140525X00011481>
- Marsh, H. W., Bormann, L., Mutz, R., Daniel, H.-D., & O’Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multi-level approaches. *Review of Educational Research*, 79(3), 1290–1326. <https://doi.org/10.3102/0034654309334143>
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications—Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160–168. <https://doi.org/10.1037/0003-066X.63.3.160>, PubMed: 18377106
- Merriman, B. (2021). Peer review as an evolving response to organizational constraint: Evidence from sociology journals, 1952–2018. *The American Sociologist*, 52(2), 341–366. <https://doi.org/10.1007/s12108-020-09473-x>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.
- Mitroff, I. I., & Chubin, D. E. (1979). Peer review at the NSF: A dialectical policy analysis. *Social Studies of Science*, 9(2), 199–232. <https://doi.org/10.1177/030631277900900203>
- Moxham, N., & Fyfe, A. (2018). The Royal Society and the prehistory of peer review, 1665–1965. *Historical Journal*, 61(4), 863–889. <https://doi.org/10.1017/S0018246X17000334>
- Neidhardt, F. (2016). Selbststeuerung der Wissenschaft durch Peer-Review-Verfahren. In D. Simon, A. Knie, S. Hornbostel, & K. Zimmermann (Eds.), *Handbuch Wissenschaftspolitik* (pp. 261–277). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-05455-7_22
- Newman, B. (2019). Authorising geographical knowledge: The development of peer review in *The Journal of the Royal Geographical Society*, 1830–c.1880. *Journal of Historical Geography*, 64, 85–97. <https://doi.org/10.1016/j.jhg.2019.03.006>
- Paltridge, B. (2017). *The discourse of peer review. Reviewing submissions to academic journals*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-48736-0>
- Peters, D. P., & Ceci, S. J. (1982). Peer review practices of psychological journals: The fate of accepted, published articles, submitted again. *Behavioral and Brain Sciences*, 5(2), 187–255. <https://doi.org/10.1017/S0140525X00011183>
- Peterson, D., & Panofsky, A. (2020). Metascience as a scientific social movement. *SocArXiv*. <https://doi.org/10.31235/osf.io/4dsqa>
- Pontille, D., & Torny, D. (2015). From manuscript evaluation to article valuation: The changing technologies of journal peer review. *Human Studies*, 38(1), 57–79. <https://doi.org/10.1007/s10746-014-9335-z>
- Posselt, J., Hernandez, T. E., Villarreal, C. D., Rodgers, A. J., & Irwin, L. N. (2020). Evaluation and decision making in higher education. In L. W. Perna (Ed.), *Higher education: Handbook of theory and research* (Vol. 35, pp. 1–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-11743-6_8-1
- Recio-Saucedo, A., Crane, K., Meadmore, K., Fackrell, K., Church, H., Fraser, S., & Blatch-Jones, A. (2022). What works for peer review and decision-making in research funding: A realist synthesis. *Research Integrity and Peer Review*, 7(1), 2. <https://doi.org/10.1186/s41073-022-00120-2>, PubMed: 35246264
- Reinhart, M. (2012). *Soziologie und Epistemologie des Peer Review*. Nomos. <https://doi.org/10.5771/9783845239415>
- Reinhart, M. (2017, July 6). *Are theories of peer review fit for the humanities and social sciences?* 2nd International Conference on Research Evaluation in the Social Sciences and Humanities, Antwerpen (Netherlands).

- Reinhart, M., & Schendzielorz, C. (2021). Peer review procedures as practice, decision, and governance. Preliminaries to theories of peer review. *SocArXiv*. <https://doi.org/10.31235/osf.io/ybp25>
- Rennie, D. (2016). Let's make peer review scientific. *Nature*, *535*, 31–33. <https://doi.org/10.1038/535031a>, PubMed: 27383970
- Rennie, D., & Flanagan, A. (2018). Three decades of peer review congresses. *JAMA*, *319*(4), 350–353. <https://doi.org/10.1001/jama.2017.20606>, PubMed: 29362775
- RoRI. (2021). *Introducing the research on research institute [Video]*. YouTube. <https://youtu.be/9NjDCKSeok>
- Ross-Hellauer, T. (2017). What is open peer review? A systematic review [version 2; peer review: 4 approved]. *F1000Research*, *6*, 588. <https://doi.org/10.12688/f1000research.11369.2>, PubMed: 28580134
- Roumbanis, L. (2022). Disagreement and agonistic chance in peer review. *Science, Technology, & Human Values*, *47*(6), 1302–1333. <https://doi.org/10.1177/01622439211026016>
- Sabaj Meruane, O., González Vergara, C., & Pina-Stranger, A. (2016). What we still don't know about peer review. *Journal of Scholarly Publishing*, *47*(2), 180–212. <https://doi.org/10.3138/jsp.47.2.180>
- Sato, S., Gyax, P. M., Randall, J., & Schmid Mast, M. (2021). The leaky pipeline in research grant peer review and funding decisions: Challenges and future directions. *Higher Education*, *82*, 145–162. <https://doi.org/10.1007/s10734-020-00626-y>, PubMed: 33041361
- Schoon, E. W. (2022). Operationalizing legitimacy. *American Sociological Review*, *87*(3), 478–503. <https://doi.org/10.1177/00031224221081379>
- Scully, M. A. (2002). Confronting errors in the meritocracy. *Organization*, *9*(3), 396–401. <https://doi.org/10.1177/135050840293004>
- Scully, M. A. (2015). Meritocracy. In J. Buckeye, D. Elm, K. Goodpaster, K. Glac, R. James, D. Jondle, D. Koehn, C. Michaelson, & F. Wettstein (Eds.), *Wiley Encyclopedia of Management* (Vol. 2, pp. 1–2). Wiley. <https://doi.org/10.1002/9781118785317.weom020075>
- Seeber, M., Vlegels, J., Reimink, E., Marušić, A., & Pina, D. G. (2021). Does reviewing experience reduce disagreement in proposals evaluation? Insights from Marie Skłodowska-Curie and COST Actions. *Research Evaluation*, *30*(3), 349–360. <https://doi.org/10.1093/reseval/rvab011>
- Severin, A., & Chataway, J. (2021). Overburdening of peer reviewers. A multi-disciplinary and multi-stakeholder perspective on causes, effects and potential policy implications. *Learned Publishing*, *34*(4), 537–546. <https://doi.org/10.1002/leap.1392>
- Severin, A., & Egger, M. (2021). Research on research funding: An imperative for science and society. *British Journal of Sports Medicine*, *55*(12), 648–649. <https://doi.org/10.1136/bjsports-2020-103340>, PubMed: 33199357
- Shah, N. B. (2022). An overview of challenges, experiments, and computational solutions in peer review. *Communications of the ACM*, *65*(6), 76–87. <https://doi.org/10.1145/3528086>
- Shepherd, J., Frampton, G. K., Pickett, K., & Wyatt, J. C. (2018). Peer review of health research funding proposals: A systematic map and systematic review of innovations for effectiveness and efficiency. *PLOS ONE*, *13*(5), e0196914. <https://doi.org/10.1371/journal.pone.0196914>, PubMed: 29750807
- Squazzoni, F. (2021). *Opinion: Despite limitations, study offers clues to gender bias*. <https://www.the-scientist.com/news-opinion/opinion-despite-limitations-study-offers-clues-to-gender-bias-68402>
- Squazzoni, F., Ahrweiler, P., Barros, T., Bianchi, F., Birukou, A., ... Willis, M. (2020). Unlock ways to share data on peer review. *Nature*, *578*, 512–414. <https://doi.org/10.1038/d41586-020-00500-y>, PubMed: 32099126
- Squazzoni, F., Bravo, G., Farjam, M., Marusic, A., Mehmani, B., ... Grimaldo, F. (2021). Peer review and gender bias: A study on 145 scholarly journals. *Science Advances*, *7*(2), eabd0299. <https://doi.org/10.1126/sciadv.abd0299>, PubMed: 33523967
- Squazzoni, F., Brezis, E., & Marušić, A. (2017). Scientometrics of peer review. *Scientometrics*, *113*(1), 501–502. <https://doi.org/10.1007/s11192-017-2518-4>, PubMed: 29056787
- Stinchcombe, A. L., & Ofshe, R. (1969). On journal editing as a probabilistic process. *The American Sociologist*, *4*(2), 116–117.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review [version 3; peer review: 2 approved]. *F1000Research*, *6*, 1151. <https://doi.org/10.12688/f1000research.12037.3>, PubMed: 29188015
- Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research Integrity and Peer Review*, *5*(1), 6. <https://doi.org/10.1186/s41073-020-00092-1>, PubMed: 32368354
- Teplitskiy, M., Peng, H., Blasco, A., & Lakhani, K. R. (2021). Is novel research worth doing? Evidence from journal peer review. *SSRN*. <https://doi.org/10.2139/ssrn.3920711>
- Tercier, J., & Callahan, M. L. (2007). *A normative model of peer review: Qualitative assessment of manuscript reviewers' attitudes towards peer review*. <https://escholarship.org/uc/item/4p90p67x>
- Thorngate, W., Dawes, R. M., & Foddy, M. (2009). *Judging merit*. Taylor & Francis. <https://doi.org/10.4324/9780203837726>
- Traag, V. A. (2021, October 4). *How do grant application criteria influence inequalities in research funding?* Funded and Unfunded Science: Academic Inequalities and Epistemic Gaps, International Research Symposium on Science Funding, Prague (Czech Republic).
- Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, *57*(1), 375–400. <https://doi.org/10.1146/annurev.psych.57.102904.190038>, PubMed: 16318600
- van den Besselaar, P., Mom, C., Cruz-Castro, L., Sanz-Menéndez, L., Hornbostel, S., ... Husu, L. (2020). *Identifying gender bias and its causes and effects*. https://www.granted-project.eu/wp-content/uploads/2021/04/GRANteD_D2.1.pdf
- van den Besselaar, P., Sandström, U., & Schiffbaenker, H. (2018). Studying grant decision-making: A linguistic analysis of review reports. *Scientometrics*, *117*(1), 313–329. <https://doi.org/10.1007/s11192-018-2848-x>, PubMed: 30220747
- van den Brink, M., & Benschop, Y. (2012). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, *19*(4), 507–524. <https://doi.org/10.1177/1350508411414293>
- Weller, A. C. (2001). *Editorial peer review. Its strengths and weaknesses*. Information Today.
- Wood, F., & Wessely, S. (2003). Peer review of grant applications: A systematic review. In F. Godlee & T. Jefferson (Eds.), *Peer review in the health sciences* (2nd ed., pp. 14–44). BMJ Books.
- Woodward, J. (1989). Data and phenomena. *Synthese*, *79*(3), 393–472. <https://doi.org/10.1007/BF00869282>
- Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, *9*(1), 66–100. <https://doi.org/10.1007/BF01553188>