A multi-disciplinary perspective on emergent and future innovations in peer review

Jonathan P. Tennant*[1], Jonathan M. Dugan[2], Daniel Graziotin[3], Damien C. Jacques[4], François Waldner[4], Daniel Mietchen[5], Yehia Elkhatib[6], Lauren B. Collister[7], Christina K. Pikas[8], Tom Crick[9], Paola Masuzzo[10], Anthony Caravaggi[11], Devin R. Berg[12], Kyle E. Niemeyer[13], Tony Ross-Hellauer[14], Sara Mannheimer[15], Lillian Rigling[16], Daniel S. Katz[17], Bastian Greshake[18], Josmel Pacheco-Mendoza[19], Nazeefa Fatima[20], Marta Poblet[21], Marios Isaakidis[22], Dasapta Erwin Irawan[23], Sébastien Renaut[24], Christopher R. Madan[25], Lisa Matthias[26], Jesper Nørgaard Kjær[27], Daniel Paul O'Donnell[28], Cameron Neylon[29], Sarah Kearns[30], Manojkumar Selvaraju[31], and Julien Colomb[32]

[1]jon.tennant.2@gmail.com; Imperial College London, London, UK, ORCID: 0000-0001-7794-0218 (*corresponding author)
[2]ORCID: 0000-0001-8525-6221
[3]Institute of Software Technology, University of Stuttgart, Stuttgart, Germany; ORCID: 0000-0002-9107-7681
[4]Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium; ORCID: 0000-0002-9069-4143 and 0000-0002-5599-7456
[5]National Institutes of Health, Bethesda, United States of America; ORCID: 0000-0001-9488-1870
[6]School of Computing and Communications, Lancaster University, Lancaster, UK; ORCID: 0000-0003-4639-436X
[7]University Library System, University of Pittsburgh, Pittsburgh, PA, USA; ORCID: 0000-0001-5767-8486
[8]Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA; ORCID: 0000-0002-4455-8036
[9]Cardiff Metropolitan University, Cardiff, UK; ORCID: 0000-0001-5196-9389
[10]Department of Biochemistry, Ghent University, Ghent, Belgium, VIB-UGent Center for Medical Biotechnology, Ghent, Belgium; ORCID: 0000-0003-3699-1195
[11]School of Biological, Earth and Environmental Sciences, University College Cork, Distillery Field, N Mall, Cork, Ireland; ORCID: 0000-0002-1763-8970
[12]Engineering & Technology Department, University of Wisconsin-Stout, Menomonie, WI, USA; ORCID: 0000-0002-1193-3848
[13]School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, OR, USA; ORCID: 0000-0003-4425-7097
[14]State and University Library, University of Goettingen, Goettingen, Germany; ORCID: 0000-0003-4470-7027
[15]Montana State University, Bozeman, Montana; ORCID: 0000-0002-1433-6782
[16]Western University Libraries, London, Ontario, Canada; ORCID: 0000-0002-9178-0422
[17]National Center for Supercomputing Applications & Department of Computer Science & Department of Electrical and Computer Engineering & School of Information Sciences, University of Illinois Urbana-Champaign, Urbana, Illinois; ORCID: 0000-0001-5934-7525
[18]Institute of Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany; ORCID: 0000-0002-9925-9623
[19]Universidad San Ignacio de Loyola, Lima, Perú; ORCID: 0000-0002-2251-8092
[20]School of Applied Sciences, University of Huddersfield, Huddersfield, UK; ORCID: 0000-0001-7791-4984
[21]Graduate School of Business and Law, RMIT University, Melbourne, Australia; ORCID: 0000-0002-0026989X
[22]Department of Computer Science, University College London, London, UK; ORCID: 0000-0001-8498-1309
[23]Department of Groundwater Engineering, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Indonesia; ORCID: 0000-0002-1526-0863
[24]Département de Sciences Biologiques, Institut de Recherche en Biologie Végétale, Université de Montréal, Montreal, QC, Canada; ORCID: 0000-0002-8111-5082
[25]Department of Psychology, Boston College, Chestnut Hill, MA, USA; ORCID: 0000-0003-3228-6501
[26]ORCID: 0000-0002-2612-2132
[27]Psychiatric Research Academy, Department of Affective Disorders, Aarhus University Hospital, Risskov, Denmark; ORCID: 0000-0001-9183-9861
[28]Department of English and Centre for the Study of Scholarly Communications, University of Lethbridge, Lethbridge AB, Canada; ORCID: 0000-0002-0127-4893
[29]Centre for Culture and Technology, Curtin University, Perth, Australia; ORCID: 0000-0002-0068-716X
[30]Department of Chemical Biology, University of Michigan, Ann Arbor, Michigan; ORCID: 0000-0001-5640-909X
[31]Integrated Gulf Biosystems, Riyadh, Saudi Arabia; Saudi Human Genome Program, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia; ORCID: 0000-0001-6439-5652
[32]Independent scholar, ORCID: 0000-0002-3127-5520

Abstract Peer review of research articles is a core part of our scholarly communication system. In spite of its clear importance, the status and purpose of peer review is often contested. What is its role in our modern, digital research and communications infrastructure? Does it perform to the high standards with which it is generally regarded? In addition, the historical evolution and present diversity of peer review is not appreciated by the research community. Differing peer review processes developed over several hundred years in a response to the needs of journals, scholarly societies, and research communities. Today, it is viewed as an inherent and altruistic responsibility among researchers, but also a common point of contention with commercial publishers. Studies of peer review have shown that it is prone to bias and abuse in numerous dimensions, frequently unreliable, and can fail to detect even fraudulent research. With the advent of Web technologies, we are now witnessing a phase of innovation and experimentation in our approaches to peer review, with new concepts and models emerging at the same times as the broader development of open research. These developments prompted us to examine emerging models of peer review from a range of disciplines and venues, and to ask how they might address some of the issues with our current systems of peer review. We examine the functionality of a range of social Web platforms, and compare these with the traits underlying a viable peer review system: quality control, quantified performance metrics as engagement incentives, and certification and reputation. Ideally any new systems will demonstrate that they out-perform current models while avoiding as many of the biases of existing systems as possible. We conclude that there is considerable scope for new peer review initiatives to be developed, each with their own potential issues and advantages. We also propose a novel hybrid platform model that, at least partially, resolves many of the technical and social issues associated with peer review, and can potentially disrupt the entire scholarly communication system. Success for any such development relies on reaching a critical threshold of research community engagement with both the process and the platform, and therefore cannot be achieved without a significant change of incentives in research environments. A major part of this relies on research communities recognizing that peer review is not a static or singular process, but a constantly evolving, and ideally improving, mechanism of quality assurance. This could greatly reduce friction to the cultural adoption of new models, and lead to a peer review process fit for a modern scholarly communication system.

1 Introduction

Peer review is the process in which experts are invited to assess the quality, novelty, validity, and potential impact of research by others, typically while it is in the form of a manuscript for an article, conference, or book (Spier, 2002). For the purposes of this article, we are exclusively addressing peer review in the context of manuscripts for research articles, unless specifically indicated; different forms of peer review are used in other contexts such as hiring, promotion, tenure, or awarding research grants (see, e.g., Fitzpatrick, 2011b, p. 16). Peer review comes in various flavors that result from different approaches to the relative timing of the review (with respect to article drafting, submission, or publication) and the transparency of the process (what is known to whom about submissions, authors, reviewers and reviews) (Ross-Hellauer, 2017). The criteria used for evaluation, including methodological soundness or expected impact are also important variables to consider. In spite of the diversity of the process, it is generally perceived as the gold standard that defines scholarly publishing by researchers and the wider public alike, and often deemed the primary determinant of scientific, theoretical, and empirical validity (Kronick, 1990). Consequently, peer review is a vital component at the core of research communication processes, with repercussions for the very structure of academia, which largely operates through a peer reviewed publication-based reward and incentive system (Moore et al., 2017). However, peer review is applied inconsistently both in theory and practice (Pontille and Torny, 2015) and generally lacks any form of transparency or formal standardization. As such, it remains difficult to know what we actually mean when we identify something as a "peer reviewed publication."

Traditionally, the function of peer review has been as a vetting procedure or gatekeeper to assist the distribution of limited resources—for instance, space in peer reviewed print publication venues, research time at specialized research facilities, or competitive research funds. Nowadays, it is also used to assess whether and how a given piece of research fits into the overall body of existing scholarly knowledge, and which journal research is suitable for and should appear in. This has consequences for whether the body of published research produced by an individual merits consideration for a more advanced position within academic or industrial research. With the advent of the Internet, the physical constraints on distribution are no longer present, and, at least in theory, we are now able to disseminate research content rapidly and at relatively negligible cost (Moore et al., 2017). This has led to the increasing popularity of digital-only publication venues that vet submissions based on the soundness of the research (e.g., *PLOS*, *PeerJ*). Such a flexibility in the filter function of peer review reduces, but does not eliminate, the role of peer review as a selective gatekeeper. Due to such innovations, ongoing discussions about peer review are intimately linked with contemporaneous developments in Open Access (OA) publishing and to broader changes in open research (Tennant et al., 2016).

The goal of this article is to investigate the historical evolution in the theory and application of peer review in a socio-technological context. We use this as the basis to consider how specific traits of consumer social Web platforms can be combined to create an optimized hybrid peer review model that is more efficient, democratic, and accountable than the traditional process.

1.1 The evolution of peer review

Any discussion on innovations in peer review must take into account its historical context. By understanding the history of scholarly publishing and the interwoven evolution of peer review, we recognize that neither are static entities, but in fact covary with each other, and therefore should be treated as such. By learning from historical experiences, we can also become more aware of how to shape future directions of peer review evolution and gain insight to what the process should look like in an optimal world. The actual term "peer review" only appears in the scientific press in the 1960s. Even in the 1970s, it was associated with grant review and not with evaluation and selection for publishing by default (Baldwin, 2017a). But the history of evaluation and selection processes for publication clearly predates the 1970s.

1.1.1 The early history of peer review

The origins of scholarly peer review of research articles are commonly associated with the formation of national academies in 17th-century Europe, although some have found foreshadowing of the practice (Al-Rahawi, c900; Spier, 2002). We call this period the *primordial* time of peer review (Figure 1). Biagioli (2002) described in detail the gradual differentiation of peer review from book censorship, and the role that state licensing and censorship systems played in 16th-century Europe; a period when monographs were the primary mode of communication. Several years after the Royal Society of London (1660) was established, it created its own in-house journal, *Philosophical Transactions*; around the same time, Denis de Sallo published the first issue of *Journal des Sçavans*. Both of these journals were first published in 1665. In London, Henry Oldenburg was appointed Secretary to the Royal Society and became the founding editor of *Philosophical Transactions*. Here, he took on the role of gathering, reporting, critiquing, and editing the work of others, as well as initiating the process of peer review as it is now commonly performed (Oldenburg, 1665; Manten, 1980). Due to this origin, peer review emerged as part of the social practices of gentlemanly learned societies. These social practices also included organizing meetings and arranging the publications of society members while being responsible for editorial curation, financial protection, and the assignment of individual prestige (Moxham and Fyfe, 2016). The development of these prototypical scientific journals gradually replaced the exchange of experimental
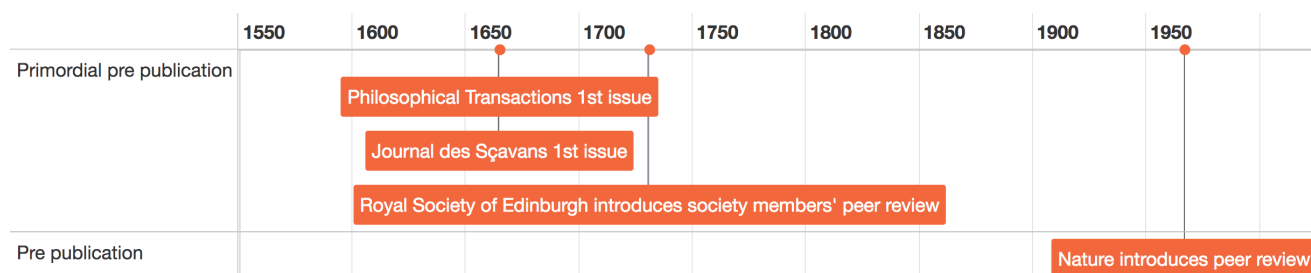
Figure 1. **A brief timeline of the evolution of peer review: The primordial times. The interactive data visualization is available at** `https://dgraziotin.shinyapps.io/peerreviewtimeline`**, and the source code and data are available at** `https://doi.org/10.6084/m9.figshare.5117260`

reports and findings through correspondence, formalizing a process that had been essentially personal and informal until then. "Peer review", during this time, was more of a civil, collegial discussion in the form of letters between authors and the publication editors (Baldwin, 2017b). Social pressures of generating new audiences for research, as well as new technological developments such as the steam-powered press, were also crucial. The purpose of developing peer reviewed journals became part of a process to deliver research to both generalist and specialist audiences, and improve the status of societies and fulfil their scholarly missions (Shuttleworth and Charnley, 2016).

From these early developments, the process of independent review of scientific reports by acknowledged experts gradually emerged. However, the review process was more similar to non-scholarly publishing, as the editors were the only ones to appraise manuscripts before printing (Burnham, 1990). As early as 1731, the Royal Society of Edinburgh adopted a formal peer review process in which materials submitted for publication in *Medical Essays and Observations* were vetted and evaluated by additional knowledgeable members (Kronick, 1990; Spier, 2002). In 1752, the United Kingdom's Royal Society created a "Committee on Papers" to review and select texts for publication in *Philosophical Transactions* (Fitzpatrick, 2011b, Chapter One). The primary purpose of this process was to select information for publication to account for the limited distribution capacity, and remained the authoritative purpose of peer review for more than two centuries.

1.1.2 Adaptation through commercialisation

Through time, the diversity, quantity, and specialization of the material presented to journal editors increased. This made it necessary to seek assistance outside the immediate group of knowledgeable reviewers from the journals' sponsoring societies (Burnham, 1990). Peer review evolved to become a largely outsourced process, which still persists in modern scholarly publishing today, where publishers call upon external specialists to validate journal submissions. The current system of peer review only became more widespread in the mid 20th century (and in some disciplines, the late 20th century or early

21st; see Graf, 2014, for an example of a major philological journal which began systematic peer review in 2011). *Nature*, now considered a top journal, did not implement such a formal peer review process until 1967 (`nature.com/nature/history/timeline_1960s.html`).

This editor-led process of peer review became increasingly important in the post-World War II decades due to the development of a modern academic prestige economy based on the perception of quality or excellence and symbolism surrounding journal-based publications (Fyfe et al., 2017; Baldwin, 2017a). The increasing professionalism of academies enabled commercial publishers to use peer review as a way of legitimizing their journals (Baldwin, 2015; Fyfe et al., 2017), and capitalized on the traditional perception of peer review as voluntary duty by academics to provide these services. A consequence of this was that peer review became a more homogenized process that enabled private publishing companies to establish a dominant, oligarchic marketplace position (Larivière et al., 2015). This represented a shift from peer review as a more synergistic activity between academics to commercial entities selling it as an added value service back to the same academic community who was performing it freely for them. The estimated cost of peer review is a minimum of $1.9bn USD per year (in 2008; (Research Information Network, 2008)), representing a substantial vested financial interest in maintaining the current process of peer review (Smith, 2010). This figure does not even account for overhead costs in publisher management or the wasteful redundancy of the reject-resubmit cycle authors enter when chasing journal prestige (Jubb, 2016), or include the time spent by typically unpaid reviewers. The result of this is that peer review has now become enormously complicated. By allowing the process of peer review to become managed by a hyper-competitive industry, developments in scholarly publishing have become strongly coupled to the transforming nature of academic research institutes. These have evolved into internationally competitive businesses that strive for quality through publisher-mediated journals by attempting to align these products with the academic ideal of research excellence (Moore et al., 2017). Such a consequence is plausibly related to, or even a consequence of, broader shifts towards a more competitive neoliberal academia and society at large. Here, emphasis is largely placed on production and

standing, value, or utility (Gupta, 2016), as opposed to the original primary focus of research on discovery and novel results.

1.1.3 The peer review revolution

In the last several decades, there have been substantial efforts to decouple peer review from the publishing process (Figure 2; Schmidt and Görögh (2017)). This has typically been done either by adopting peer review as an overlay process on top of formally published research articles or by pursuing a "publish first, filter later" protocol, with peer review taking place after the initial publication of research results (Moed, 2007; McKiernan et al., 2016). Here, the meaning of "publication" becomes "making public", as opposed to the traditional sense where it also implies peer reviewed. In fields such as Physics and Mathematics, it has traditionally been commonplace for authors to send their colleagues either paper or electronic copies of their manuscripts for pre-submission evaluation. Launched in 1991, *arXiv* (`arxiv.org`) formalized this process by creating a central network for whole communities to access such e-prints. Today, *arXiv* has more than one million e-prints from various research fields and receives more than 8,000 monthly submissions (arXiv, 2017). Here, e-prints or pre-prints are not formally peer reviewed prior to publication, but still undergo a certain degree of moderation in order to filter out non-scientific content. This practice represents a significant shift as public dissemination was decoupled from a traditional peer review process, resulting in increased visibility and citation rates (Davis and Fromerth, 2007; Moed, 2007). The launch of *Open Journal Systems* (`openjournalsystems.com`; OJS) in 2001 offered a step towards bringing journals and peer review back to their community-led roots. As of 2015, the OJS platform provided the technical infrastructure and editorial and peer review workflow management support to more than 10,000 journals (Public Knowledge Project, 2016). Its exceptionally low cost was perhaps most responsible for the preferential adoption (around half) of these journals appearing in the developing world (Edgar and Willinsky, 2010).

More recently, there has been a new wave of innovation in peer review, which we term "the revolution" phase (Figure 2; note that this is a non-exhaustive overview of the peer review landscape). The pace of this is accelerating rapidly, with the majority of changes occurring in the last five to ten years. This could be related to initiatives such as the San Francisco Declaration on Research Assessment (`ascb.org/dora/`; DORA), that called for systemic changes in the way that scientific research outputs are evaluated. Digital-born journals, such as *PLOS ONE*, introduced commenting on published papers. This spurred developments in cross-publisher annotation platforms like *PubPeer* and *PaperHive*. Some journals, such as *F1000 Research* and *The Winnower*, rely exclusively on a model where peer review is conducted after the manuscripts are made publicly available. Other ser-

vices, such as *Publons*, enable reviewers to claim recognition for their activities as referees. Platforms such as *ScienceOpen* provide a search engine combined with peer review across publishers on all documents, regardless of whether manuscripts have been previously reviewed. Each of these innovations has partial parallels to other social Web applications or platforms in terms of transparency, reputation, performance assessment, and community engagement. It remains to be seen whether these innovations and new models of evaluation will become more popular than traditional peer review.

1.2 The role and purpose of modern peer review

Due to the increasingly systematic use of external peer review, its processes have become entwined with the core activities of scholarly communication. Without approval through peer review to assess importance, validity, and journal suitability, research articles will not be sent to print. The historical motivation for selecting amongst submitted articles or distribution was primarily economic. With scholarly publishing turning into an essentially loss-making business, the costs of printing and paper needed to be limited (Fyfe, 2015). The rising number of submissions, particularly in the 20th century, required distributing the management of this selection process. While in the digital world the costs of dissemination have dropped, the marginal cost of publishing articles is far from zero (e.g., due to time and management, hosting, marketing, technical and ethical checks, among other services). The economic motivations for still imposing selectivity in a digital environment, and applying peer review as a mechanism for this, have received limited attention or questioning, and is often regarded as just how things are done. Selectivity is now often attributed to quality control, but is based on the false assumption that peer review requires careful selection of specific reviewers to assure a definitive level of adequate quality, termed the "Fallacy of Misplaced Focus" by Kelty et al. (2008).

In many cases, there is an attempt to link the goals of peer review processes with Mertonian norms (Merton, 1973; Lee et al., 2013) (i.e., universalism, communalism, disinterestedness, and organized skepticism) as a way of showing their relation to shared community values. The Mertonian norm of organized skepticism is the most obvious link, while the norm of disinterestedness can be linked to efforts to reduce systemic bias, and communalism to the expectation of contribution to peer review as part of community membership (i.e., duty). In contrast to the emphasis on supposedly shared social values, relatively little attention has been paid to the diversity of processes of peer review across journals, disciplines, and time. This is especially the case as the (scientific) scholarly community appears overall to have a strong investment in a "creation myth" that links the beginning of scholarly publishing—the founding of *The Philosophical Transactions of the Royal Society*—to the
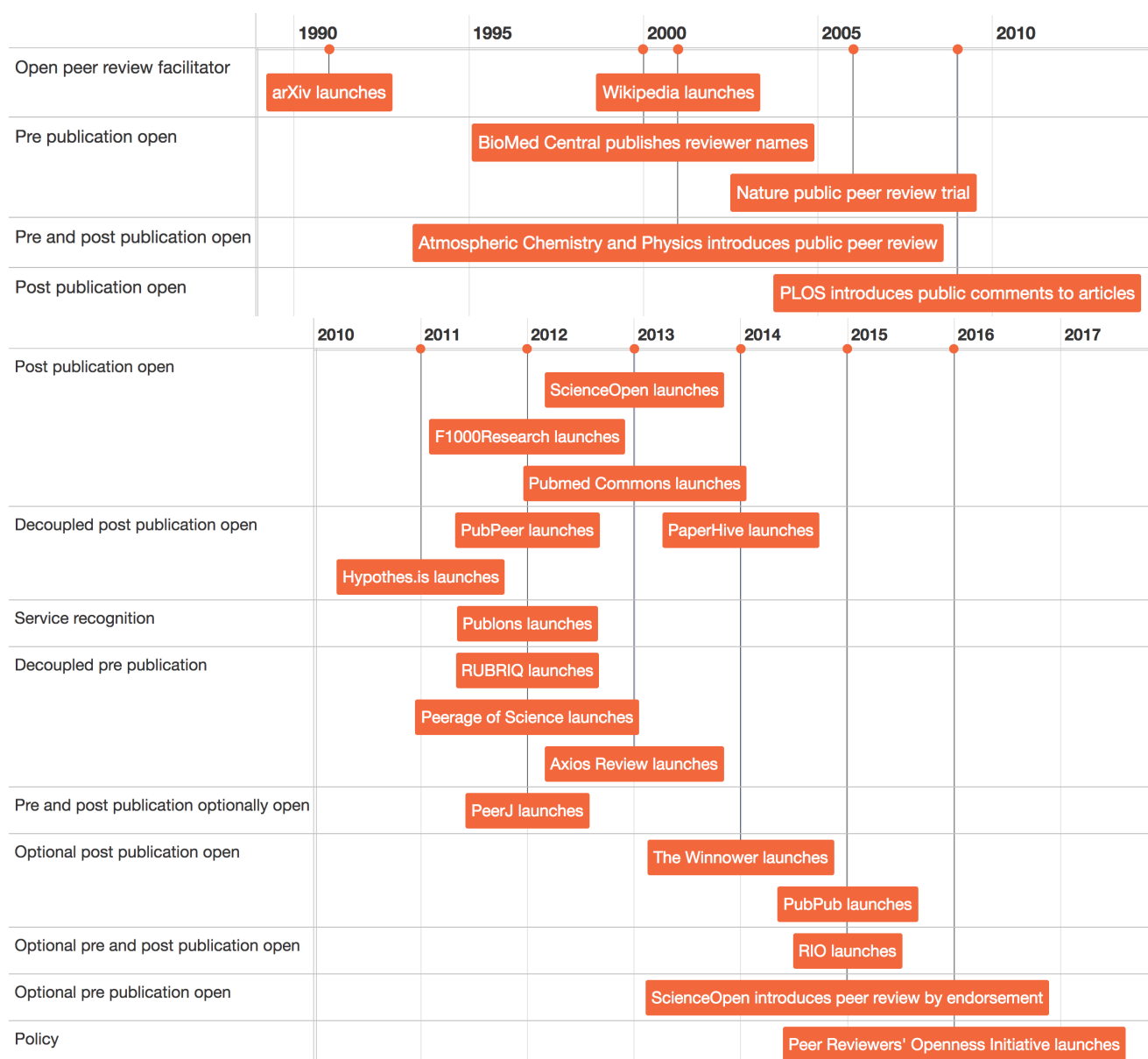
Figure 2. A brief timeline of the evolution of peer review: The revolution.

invention of peer review. The two are often regarded to be coupled by necessity, largely ignoring the complex and interwoven history of peer review and publishing. This has consequences, as the individual identity as a scholar is strongly tied to specific forms of publication that are evaluated in particular ways (Moore et al., 2017). A scholar's first research article, PhD thesis, or first book are significant life events. Membership of a community, therefore, is validated by the peers who review this newly contributed work. Community investment in the idea that these processes have "always been followed" appears very strong, but ultimately remains a fallacy.

As mentioned above, there is an increasing quantity and quality of research that examines how publication processes, selection, and peer review evolved from the 17th to the early 20th century, and how this relates to broader social patterns (Baldwin, 2017a,b; Moxham

and Fyfe, 2016). However, there is much less research critically exploring the diversity of selection and peer review processes in the mid- to late-20th century. Indeed, there seems to be a remarkable discrepancy between the historical work we do have (Shuttleworth and Charnley, 2016; Gupta, 2016; Baldwin, 2017a) and apparent community views that "we have always done it this way," alongside what sometimes feels like a wilful effort to ignore the current diversity of practice.

Such a discrepancy between a dynamic history and re-membered consistency could be a consequence of peer review processes being central to both scholarly identity as a whole and to the identity and boundaries of specific communities (Moore et al., 2017). Indeed, this story link-ing identity to peer review is taught to junior researchers as a community norm, often without the much-needed historical context. More work on how peer review, along-

side other community practices, contributes to community building and sustainability would be valuable. Examining criticisms of conventional peer review and proposals for change through the lens of community formation and identity may be a productive avenue for future research.

1.3 Criticisms of the conventional peer review system

In spite of its clear relevance, widespread acceptance, and long-standing practice, the academic community does not appear to have a clear consensus on the operational functionality of peer review, and what its effects in a diverse modern research world are. There is a discrepancy between how peer review is regarded as a process, and how it is actually performed. While peer review is still generally perceived as key to quality control for research, others have begun to note that mistakes are becoming ever more frequent in the process (Smith, 2006; Margalida and Colomer, 2016), or at least that peer review is problematic and not being applied as rigorously as generally perceived (Eckberg, 1991; Kostoff and N, 1995; Jefferson et al., 2002; Schroter et al., 2006; Cole, 2000; Ghosh et al., 2012; Ross-Hellauer, 2017; Walker and Rocha da Silva, 2015). One consequence of this is that COPE, the Committee on Publication Ethics (`publicationethics.org`), was established in 1997 to address potential cases of abuse and misconduct during the publication process. Yet, the effectiveness of this initiative at a system-level remains unclear. A popular editorial in *The BMJ* stated that peer review is "slow, expensive, profligate of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless at detecting fraud," with evidence supporting each of these quite serious allegations (Smith, 2006). However, beyond editorials, there now exists a substantial corpus of studies that critically examine the technical aspects of peer review which, taken together, should be extremely worrisome. This is especially so given that traditional peer review is still viewed almost dogmatically as a gold standard for the publication of research results, and as the process which mediates knowledge dissemination to the public.

The issue is that, ultimately, this uncertainty in standards and implementation can potentially lead to, or at least be viewed as the cause of, widespread failures in research quality and integrity (Jefferson et al., 2002; Ioannidis, 2005) and even the rise of formal retractions in extreme cases (Steen et al., 2013). Issues resulting from peer review failure range from simple gate-keeping errors, based on differences in opinion of the perceived impact of research, to failing to detect fraudulent or incorrect work, which then enters the scientific record (Moore et al., 2017; Pocock et al., 1987; Gøtzsche, 1989; Baxt et al., 1998; Smith, 2006; Haug, 2015; Schroter et al., 2004). A final issue regards peer review by and for non-native English speaking authors, which can lead to cases of linguistic inequality and language-oriented research segregation, in a world where research is increasingly be-

coming more globally competitive (Salager-Meyer, 2008, 2014). All of this suggests that, while the idea of peer review remains logical, it is the implementation of it that requires attention.

1.3.1 Peer review needs to be peer reviewed

Attempts to reproduce how peer review selects what is worthy of publication demonstrate that the process is generally adequate for detecting reliable research, but often fails to recognize the research that has the greatest impact (Siler et al., 2015; Moore et al., 2017; Mahoney, 1977). Many now regard the traditional peer review model as sub-optimal in that it causes publication delays, impacting the communication of novel research (Bornmann and Daniel, 2010; Vines, 2015b; Brembs, 2015; Eisen, 2011; Jubb, 2016). Reviewer fatigue (Breuning et al., 2015) and redundancy when articles go through multiple rounds of peer review at different journal venues (Moore et al., 2017; Jubb, 2016) are just some of the criticisms levied at the technical implementation of peer review. In addition, some view traditional peer review as flawed because it operates within a closed and opaque system. This makes it impossible to trace the discussions that led to (sometimes substantial) revisions to the original research (Bedeian, 2003), as well as the decision process leading to the final publication.

On top of all of these potential issues, some critics go even further in stating that, at its worst, peer review can be seen as detrimental to research. By operating as a closed system, it protects the status quo and suppresses research viewed as radical, innovative, or contrary to the theoretical perspectives of referees (Merton, 1968; Mahoney, 1977; Horrobin, 1990; Alvesson and Sandberg, 2014; Benda and Engels, 2011), even though it is precisely these factors that underpin and advance research. As a consequence, questions to the competency and integrity of traditional peer review arise, such as: who are the gatekeepers and how are their gates constructed; what is the balance between author-reviewer-editor tensions; what are the inherent biases associated with this; does this enable a fair or structurally inclined system of peer review to exist; and what are the repercussions for this on our knowledge generation and communication systems?

In spite of all of these criticisms, it remains clear that the ideal of peer review still plays a fundamental role in scholarly communication (Goodman et al., 1994; Pierie et al., 1996; Ware, 2008) and retains a high level of respect from the research community (Bedeian, 2003; Greaves et al., 2006; Gibson et al., 2008). One primary reason why peer review has persisted is that it remains a unique way of differentiating research publications from other types of literature, including blogs, media articles, and books, and assigning academic credit to the former. This perception, combined with a general lack of awareness or appreciation of the historic context of peer review, research examining its potential flaws, and the conflation of the process with the ideology, has sustained

its ubiquitous usage and continued proliferation in academia. This has led to the widely-held perception that peer review is a singular and static process, and to its acceptance as a social norm. This in turn defines the difficulty in moving away from a process that has now become so deeply embedded within oligarchic research institutes. The consequence of this is that, irrespective of any systemic flaws, peer review remains one of the essential pillars of trust when it comes to scientific communication (Haider and Åström, 2017).

In this article, we summarize the ebb and flow of the debate around the various and complex aspects of conventional (editorially-controlled) peer review. In particular, we highlight how innovative systems are attempting to resolve the major issues associated with traditional models, explore how new platforms could improve the process in the future, and consider what this means for the identity, role, and purpose of peer review within diverse research communities. The aim of this discussion is not to undermine any specific model of peer review in a quest for systemic upheaval, or to advocate any particular alternative model. Rather, we acknowledge that the idea of peer review is critical for research and advancing our knowledge, and as such we provide a foundation here for future exploration and creativity in diversifying and improving an essential component of scholarly communication.

2 The traits and trends affecting modern peer review

Over time, three principal forms of journal peer review have evolved: single-blind, double-blind, and open (Table 1). Of these, single-blind, where reviewers are anonymous but authors are not, is the most widely-used in most disciplines because the process is comparably less onerous and less expensive to operate than the alternatives. Double-blind peer review, where both authors and reviewers are reciprocally anonymous, requires considerable effort to remove all traces of the author's identity from the manuscript under review (Blank, 1991). For a detailed comparison of double versus single blind review, Snodgrass (2007) provides an excellent summary. These are generally considered to be the traditional forms of peer review, with the advent of open peer review introducing substantial additional complexity into the discussion (Ross-Hellauer, 2017).

Table 1. Types of peer review.

| | | Author Identity | |
		Hidden	Known
Reviewer Identity	Hidden	Double blind	Single blind
	Known	–	Open

The diversification of peer review is intrinsically coupled

with wider developments in scholarly publishing. When it comes to the gate-keeping function of peer review, innovation is noticeable in some digital-only, or "born open," journals, such as *PLOS ONE* and *PeerJ*. These explicitly request referees to ignore any notion of novelty, significance, or impact, before it becomes accessible to the research community. Instead, reviewers are asked to focus on whether the research was conducted properly and that the conclusions are based on the presented results. This arguably more objective method has met some resistance, even receiving the somewhat derogatory term "peer review lite" from some corners of the scholarly publishing industry (Pinfield, 2016). Such a perception is largely a hangover from the commercial age of publishing, and now seems superfluous and discordant with any modern Web-based model of scholarly communication. The relative timing of peer review to publication is a further major innovation, with journals such as *F1000 Research* publishing prior to any formal peer review process. Some of the advantages and disadvantages of these different variations of open peer review are explored in Table 2.

2.1 The development of open peer review

Novel ideas about "Open Peer Review" (OPR) systems are rapidly emerging, and innovation has been accelerating over the last several years (Figure 2; Table 3). The advent of OPR is complex and often multiple aspects of peer review are used inter-changeably or are conflated without appropriate prior definition. Currently, there is no formally established definition of OPR that is accepted by the scholarly research and publishing community (Ford, 2013). The most simple definitions by McCormack (2009) and Mulligan et al. (2008) presented OPR as a process that does not attempt "to mask the identity of authors or reviewers" (McCormack, 2009, p.63), thereby explicitly referring to open in terms of personal identification or anonymity. Ware (2011, p.25) expanded on reviewer disclosure practices: "Open peer reviews can mean the opposite of double blind, in which authors' and reviewers' identities are both known to each other (and sometimes publicly disclosed), but discussion is complicated by the fact that it is also used to describe other approaches such as where the reviewers remain anonymous but their reports are published." Other authors define OPR distinctly, such as including the publication of all dialogue during the process (Shotton, 2012), or running it as a publicly participative commentary (Greaves et al., 2006). A recent survey by *OpenAIRE* found 122 different definitions of OPR in use, exemplifying the extent of this issue. This diversity was distilled into a single proposed definition comprising seven different open traits: participation, identity, reports, interaction, platforms, pre-review manuscripts, and final-version commenting (Ross-Hellauer, 2017).

A core question is how to transform traditional peer review into a process aligned with the latest advances in what is now widely termed "open science". This is

Table 2. Advantages and disadvantages of the different approaches to peer review (NPRC, Neuroscience Peer Review Consortium)

Type	Description	Pros/Benefits	Cons/Risks	Examples
Pre-peer review commenting	Informal commenting and discussion on a publicly available pre-publication manuscript draft (i.e., preprints)	Rapid, transparent, public, relatively low cost (free for authors), open commenting	Variable uptake, fear of scooping, fear of journal rejection, fear of premature communication	bioRxiv, SocArXiv, engrXiv, PeerJ pre-prints, Figshare, Zenodo
Pre-publication	Formal and editorially-invited evaluation of a piece of research by selected experts in the relevant field	Editorial moderation, provides at least some consistent form of quality control for all published work	Non-transparent, impossible to evaluate, biased, secretive, exclusive, unclear who "owns" reviews	Nature, Science, New England Journal of Medicine, Cell, The Lancet
Post-publication	Formal and optionally-invited evaluation of research by selected experts in the relevant field, subsequent to publication	Rapid publication of research, public, transparent, can be editorially-moderated	Filtering of "bad research" occurs after publication, relatively low uptake	F1000Research, ScienceOpen, Research Ideas and Outcomes (RIO), The Winnower
Post-publication commenting	Informal discussion of published research, independent of any formal peer review that may have already occurred	Can be performed on third-party platforms, anyone can contribute, public	Comments can be rude or of low quality, comments across multiple platforms lack inter-operability, low visibility, low uptake	PubMed Commons, PeerJ, PLOS ONE, ScienceOpen
Collaborative	Referees, and often editors, participate in the assessment of scientific manuscripts through interactive comments to reach a consensus decision and a single set of revisions and comments	Iterative, editors sign reports, can be integrated with formal process, deters low quality submissions	Can be additionally time-consuming, discussion quality variable, peer pressure and influence can tilt the balance	eLife, Frontiers, Atmospheric Chemistry and Physics
Portable	Authors can take referee reports to multiple consecutive venues, often administered by a third-party service	Reduces redundancy or duplication, saves time	Low uptake by authors, low acceptance by journals, high cost	BioMed Central journals, NPRC, Rubriq, Peerage of Science, MECA
Recommendation services	Post-publication evaluation and recommendation of significant articles, often through a peer-nominated consortium	Crowd-sourced literature discovery, time saving, "prestige" factor when inside a consortium	Paid services (subscription only), time consuming on recommender side, exclusive	F1000Prime, CiteULike, ScienceOpen
De-coupled post-publication (annotation services)	Comments or highlights added directly to highlighted sections of the work. Added notes can be private or public	Rapid, crowd-sourced and collaborative, cross-publisher, low threshold for entry	Non-interoperable, multiple venues, effort duplication, relatively unused, genuine critiques reserved	PubPeer, Hypothesis, PaperHive, PeerLibrary

tied to broader developments in how we as a society communicate, thanks to the inherent capacity that the Web provides for open, collaborative, and social communication. Many of the suggestions and new models for improving peer review are geared towards increasing the transparency and ultimately the reliability, efficiency, and accountability of the publishing process, and aligning peer review norms to support these aims. These traits are desired by all actors in the system, and increasing transparency moves peer review towards a more open model.

However, the context of this transparency and the implications of different levels of transparency at different stages of the review process are both very rarely explored, and achieving transparency is difficult at a variety of levels. How and where we inject transparency into the system has implications for the magnitude of transformation and, therefore, the general concept of OPR is highly heterogeneous in meaning, scope, and consequences. New suggestions to modify peer review vary between fairly incremental small-scale changes

to those that encompass an almost total and radical transformation of the present system. The various parts of the "revolutionary" phase of peer review undoubtedly have different combinations of these OPR traits, and within this remains a very heterogeneous landscape. Table 3 provides an overview of the advantages and disadvantages of the different approaches to anonymity and openness in peer review.

In this article, we regard OPR as a process fulfilling any of the following three primary criteria:

1. Referee names are identified to the authors and the readership;

2. Referee reports are made publicly available under an open license;

3. Peer review is not restricted to invited referees only.

With all of these complex evolutionary trajectories, it is clear that peer review is undergoing a phase of experimentation in line with the evolving scholarly

ecosystem. However, despite the range of new innovations, the engagement with these experimental open models is still far from common. The entrenchment of the ubiquitously practiced and much more favored traditional model (which, as noted above, is also diverse) is ironically non-traditional but nonetheless currently revered. Practices such as self-publishing and predatory or deceptive publishing cast a shadow of doubt on the validity of research posted openly online that follow these models, including those with traditional scholarly imprints (Fitzpatrick, 2011a; Tennant et al., 2016). The inertia hindering widespread adoption of new models of peer review can be ascribed to what is often termed "cultural inertia" within scholarly research. Cultural inertia, the tendency of communities to cling to a traditional trajectory, is shaped by a complex ecosystem of individuals and groups, who often have highly polarized motivations (i.e., capitalistic commercialism versus knowledge generation versus careerism versus output measurement), and an academic hierarchy that imposes a power dynamic that can suppress innovative practices (Burris, 2004; Magee and Galinsky, 2008).

The ongoing discussions and innovations around peer review (and OPR) can be sorted into four main categories, which are examined in more detail below. Each of these feed into the wider issues of incentivizing engagement, providing appropriate recognition and certification, and quality control and moderation:

1. How can referees receive credit or recognition for their work, and what form should this take;

2. Should referee reports be published alongside manuscripts;

3. Should referees remain anonymous or have their identities disclosed;

4. Should peer review occur prior or subsequent to the publication process (i.e., publish then filter).

2.2   Giving credit to peer reviewers

A vast majority of researchers see peer review as an integral and fundamental part of their work. They often even consider peer review to be part of an altruistic cultural duty or a quid pro quo service, closely associated with the identity of being part of their research community. Generally, journals do not provide any remuneration or compensation for these services. Notable exceptions are the UK-based publisher *Veruscript* (`veruscript.com/about/who-we-are`) and *Collabra* (`collabra.org/about/our-model`), published by University of California Press. To be invited to review a research article is perceived as a great honor, especially for junior researchers, due to the recognition of expertise—i.e., the attainment of the level of a peer. However, the current system is facing new challenges as the number of published papers continues to increase rapidly (Albert et al., 2016), with more than one million articles published in peer reviewed,

English-language journals every year (Larsen and Von Ins, 2010). Some estimates are even as high as 2–2.5 million per year (Plume and van Weijen, 2014), and this number is expected to double approximately every nine years at current rates (Bornmann and Mutz, 2015). There are several possible solutions to this issue:

- Increase the total pool of potential referees,

- Increase acceptance rates to avoid review duplication,

- Decrease the number of referees per paper, and/or

- Decrease the time spent on peer review.

Of these, the latter two can both potentially reduce the quality of peer review, open or otherwise, and therefore affect the overall quality of published research. Paradoxically, as the Internet empowers us to communicate information virtually instantaneously, the turn around time for peer reviewed publications is as far from this as it ever has been. One potential solution to this is to encourage referees by providing additional recognition and credit for their work. The present lack of bona fide incentives for referees is perhaps the main factor responsible for indifference to editorial outcomes, which ultimately leads to the increased proliferation of low quality research (D'Andrea and O'Dwyer, 2017).

2.2.1   Traditional methods of recognition

One current way to recognize peer review is to thank anonymous referees in the Acknowledgement sections of published papers. In these cases, the referees will not receive any public recognition for their work, unless they explicitly agree to sign their reviews. Another common form of acknowledgement is a private thank you note from the journal or editor, which usually takes the form of an automated email upon completion of the review. In addition, journals often list and thank all reviewers in a special issue or on their website once a year, thus providing another way to credit reviewers. Another idea that journals and publishers have tried implementing is to list the best reviewers for their journal (e.g., by Vines (2015a) for *Molecular Ecology*), or, on the basis of a suggestion by Pullum (1984), naming referees who recommend acceptance in the article colophon (a single blind version of this recommendation was adopted by *Digital Medievalist* from 2005-2016; see Wikipedia contributors, 2017, and bit.ly/DigitalMedievalistArchive for examples preserved in the Internet Archive. *DM* stopped using this model and removed the colophon as part of its move to the Open Library of Humanities; cf. journal.digitalmedievalist.org). As such, authors can then integrate this into their scholarly profiles in order to differentiate themselves from other researchers or referees. Currently, most tenure and review committees do not consider peer review activities as required or sufficient in the process of professional advancement or tenure evaluation. Instead, it is viewed as expected or normal behaviour for all researchers to contribute in some form to peer review.

Table 3. Pros and cons of different approaches to anonymity in peer review

Approach	Description	Pros/Benefits	Cons/Risks	Examples
Single-blind peer review	Referees are not revealed to the authors, but referees are aware of author identities	Allows reviewers to view full context of an author's other work, detection of COIs, more efficient	Prone to bias, authors not protected, exclusive, non-verifiable, referees can often be identified anyway	Most biomedical and physics journals, PLOS ONE, Science
Double-blind peer review	Authors and the referees are reciprocally anonymous	Increased author diversity in published literature, protects authors and reviewers from bias, more objective	Still prone to abuse and bias, secretive, exclusive, non-verifiable, referees can often be identified anyway, time consuming	Nature, most social sciences journals
Triple-blind peer review	Authors and their affiliations are reciprocally anonymous to handling editors and reviewers	Eliminates geographical, institutional, personal and gender biases, work evaluated based on merit	Incompatible with pre-prints, low-uptake, non-verifiable, secretive	Science Matters
Private, open peer review	Referee names are revealed to the authors pre-publication, if the referees agree, either through an opt-in or opt-out mechanism	Protects referees, no fear of reprisal for critical reviews	Increases decline to review rates, non-verifiable	PLOS Medicine, Learned Publishing
Unattributed peer review	If referees agree, their reports are made public but anonymous when the work is published	Reports publicized for context and re-use	Prone to abuse and bias, secretive, exclusive, non-verifiable	EMBO Journal
Optional open peer review	As single-blind peer review, except that the referees are given the option to make their review and their name public	Increased transparency	Gives an unclear pictures of the review process if not all reviews are made public	PeerJ, Nature Communications
Pre-publication open peer review	Referees are identified to authors pre-publication, and if the article is published, the full peer review history together with the names of the associated referees is made public	Transparency, increased integrity of reviews	Fear: referees may decline to review, or be unwilling to come across too critically or positively	The medical BMC-series journals, The BMJ
Post-publication open peer review	The referee reports and the names of the referees are always made public regardless of the outcome of their review	Fast publication, transparent process	Fear: referees may decline to review, or be unwilling to come across too critically or positively	F1000Research, ScienceOpen, PubPub
Peer review by endorsement (PRE)	Pre-arranged and invited, with referees providing a "stamp of approval" on publications	Transparent, cost-effective, rapid, accountable	Low uptake, prone to selection bias, not viewed as credible	RIO Journal, ScienceOpen

2.2.2 Increasing demand for recognition

Traditional approaches of credit fall short of any sort of systematic feedback or recognition, such as that granted through publications. A change here is clearly required for the wealth of currently unrewarded time and effort given to peer review by academics. A recent survey of nearly 3,000 peer reviewers by the large commercial publisher *Wiley* showed that feedback and acknowledgement for work as referees are valued far above either cash reimbursements or payment in kind (Warne, 2016). As of today, peer review is poorly acknowledged by practically all research assessment bodies, institutions, granting agencies, as well as publishers. *Wiley's* survey reports that 80% of researchers agree that there is insufficient recognition for peer review as a valuable research activity and that researchers would actually commit more time to peer review if it became a formally recognized activity for assessments, funding opportunities, and promotion (Warne, 2016). While this may be true, it is important to note that commercial publishers, including *Wiley*, have a vested interest in retaining the current, freely provided service of peer review since this is what provides their journals the

main stamp of legitimacy and quality ("added value") as society-led journals. Therefore, one of the root causes for the lack of appropriate recognition and incentivization is, ironically, publishers themselves, who have strong motivations to find non-monetary forms of reviewer recognition. Indeed, the business model of almost every large scholarly publisher is predicated on free work by peer reviewers, and it is unlikely that the present system would function financially with market-rate reimbursement of peer reviewers. Hence, this survey could represent a biased view of the actual situation. Other research shows a similar picture, with approximately 70% of respondents to a small survey done by Nicholson and Alperin (2016) indicating that they would list peer review as a professional service on their curriculum vitae. 27% of respondents mentioned formal recognition in assessment as a factor that would motivate them to participate in public peer review. These numbers indicate that the lack of credit referees receive for peer review is a contributing factor to the perceived stagnation of the traditional models. Furthermore, acceptance rates are lower in humanities and social sciences, and higher in physical sciences and engineering

journals (Ware, 2008). This means there are distinct disciplinary variations in the number of reviews performed by a researcher relative to their publications and suggests that there is scope for using this to either provide different incentive structures or to increase acceptance rates and therefore decrease referee fatigue (Lyman, 2013).

2.2.3 Progress in crediting peer review

Any acknowledgement model to credit reviewers also raises the obvious question of how to facilitate this model within an anonymous peer review system. By incentivizing peer review, much of its potential burden can be alleviated by widening the potential referee pool. This can also help to diversify the process and inject transparency into peer review, a solution that is especially appealing when considering that it is often a small minority of researchers who perform the vast majority of peer reviews (Gropp et al., 2017; Fox et al., 2017); for example, in biomedical research, only 20 percent of researchers perform 70–95 percent of the reviews (Kovanis et al., 2016). In 2014, a working group on peer review services (CASRAI) was established to "develop recommendations for data fields, descriptors, persistence, resolution, and citation, and describe options for linking peer-review activities with a person identifier such as *ORCID*" (Paglione and Lawrence, 2015). The idea here is that by being able to standardize peer review activities, it becomes easier to describe, attribute, and therefore recognize and reward them.

The *Publons* platform provides a semi-automated mechanism to formally recognize the role of editors and referees who can receive due credit for their work as referees, both pre- and post-publication. Researchers can also choose if they want to publish their full reports depending on publisher and journal policies. *Publons* also provides a ranking for the quality of the reviewed research article, and users can endorse, follow, and recommend reviews. Other platforms, such as *F1000 Research* and *ScienceOpen*, link post-publication peer review activities with *CrossRef* DOIs to make them more citable, essentially treating them equivalent to a normal Open Access research paper. *ORCID* (Open Researcher and Contributor ID) provides a stable means of integrating with platforms such as *Publons* and *ImpactStory* in order to receive due credit for reviews. *ORCID* is rapidly becoming part of the critical infrastructure for OPR, and greater shifts towards open scholarship (Dappert et al., 2017). Exposing peer reviews through these platforms links accountability to receiving credit. Therefore, they offer possible solutions to the dual issues of rigor and reward, while potentially ameliorating the growing threat of reviewer fatigue. Whether such initiatives will be successful remains to be seen, although *Publons* was recently acquired by *Clarivate Analytics*, suggesting that the process could become commercialized as this domain rapidly evolves (Van Noorden, 2017). In spite of this, the outcome is most likely to be dependent on whether funding agencies and those in charge of tenure, hiring, and promotion will use peer review activities to help evaluate candidates. This is likely dependent on whether research communities themselves choose to embrace any such crediting or accounting systems for peer review.

2.3 Publishing peer review reports

The rationale behind publishing referee reports lies in providing increased context and transparency to the peer review process—the making of the sausage, so to speak. Often, valuable insights are shared in reviews that would otherwise remain hidden if not published. By publishing reports, peer review then has the potential to become a supportive and collaborative process that is viewed more as an ongoing dialogue between groups of scientists to progressively assess the quality of research. Furthermore, the reviews themselves are opened up for analysis and inspection, including how authors respond to reviews themselves, which adds an additional layer of quality control and a means for accountability and verification. There are additional educational benefits to publishing peer reviews, such as training purposes or for journal clubs. At the present, some publisher policies are extremely vague about the re-use rights and ownership of peer review reports (Schiermeier, 2017).

In a study of two journals, one where reports were not published and another where they were, Bornmann et al. (2012) found that publicized comments were much longer. Furthermore, there was an increased chance that they may result in a constructive dialogue between the author, reviewers, and wider community, and might therefore be better for improving the content of a manuscript. On the other hand, unpublished reviews tend to have had more of a selective function to determine whether a manuscript is appropriate for a particular journal (i.e., focusing on the editorial process). Therefore, depending on the journal, different types of peer review could be better suited to perform different functions, and therefore optimized in that direction. Transparency of the peer review process can also be used as an indicator for peer review quality, thereby potentially enabling the tool to predict quality in new journals in which the peer review model is known (Godlee, 2002; Morrison, 2006; Wicherts, 2016), if desired. Journals with higher transparency ratings, moreover, were less likely to accept flawed papers and showed a higher impact as measured by Google Scholar's h5-index.

It is ironic that, while assessments of articles can never be evidence-based without the publication of referee reports, they are still almost ubiquitously regarded as having an authoritative stamp of quality. The issue here is the attainment of peer reviewed status will always be based on an undefined, and only ever relative, quality threshold due to the opacity of the process. This is quite an unscientific practice, and instead, researchers rely almost entirely on heuristics and trust for a concealed process and the intrinsic reputation of the journal, rather than anything legitimate. This can ultimately result

in what is termed the "Fallacy of Misplaced Finality" by Kelty et al. (2008), describing the assumption that research has a single, final form, to which everyone applies different criteria of quality.

Publishing peer review reports appears to have little or no impact on the overall process but may encourage more civility from referees. In a small survey, Nicholson and Alperin (2016) found that approximately 75% of survey respondents (n=79) perceived that public peer review would change the tone or content of the reviews, and 80% of responses indicated that performing peer reviews that would be eventually be publicized would not require a significantly higher amount of work. However, the responses also indicated that an incentive is needed for referees to engage in open peer review. This would include recognition by performance review or tenure committees (27%), peers publishing their reviews (26%), being paid in some way such as with an honorarium or waived APC (24%), and getting positive feedback on reviews from journal editors (16%). Only 3% (one response) indicated that nothing could motivate them to participate in an open peer review of this kind. Leek et al. (2011) showed that when referees' comments were made public, significantly more cooperative interactions were formed, while the risk of incorrect comments decreased. Moreover, referees and authors who participated in cooperative interactions had a reviewing accuracy rate that was 11% higher. On the other hand, the possibility of publishing the reviews online has also been associated with a high decline rate among potential peer reviewers and an increase in the amount of time taken to write a review, but with no effect on review quality (van Rooyen et al., 2010). This suggests that the barriers to publishing review reports are inherently social, rather than technical.

When *BioMed Central* launched in 2000, it quickly recognized the value in including both the reviewers' names and the peer review history (pre-publication) alongside published manuscripts in their medical journals. Since then, further reflections on open peer review (Godlee, 2002) led to the adoption of a variety of OPR models. For example, the *Frontiers* series now publishes all referee names alongside articles, *EMBO* journals publish a review process file with the articles, with referees remaining anonymous but editors being named, and *PLOS* added public commenting features to articles they published in 2009. More recently launched journals such as *PeerJ* have a system where both the reviews and the names of the referees can optionally be made public, and journals such as *Nature Communications* and the *European Journal of Neuroscience* have recently started to adopt this method of OPR as well.

Unresolved issues with posting review reports include whether or not it should be conducted for ultimately unpublished manuscripts, the impact of author identification or anonymity, and if the announcement of author's career stage has potential consequences on their reputations. Furthermore, the actual readership and usage of published reports remains ambiguous in a world where researchers are typically already inundated with published articles to read. The benefits of publicizing reports might not be seen until further down the line from the initial publication and, therefore, their immediate value might be difficult to convey and measure in current research environments. Finally, different populations of reviewers with different cultural norms and identities will undoubtedly have varying perspectives on this issue, and it is unlikely that any single policy or solution to posting referee reports will ever be widely adopted.

2.4 Eponymous versus anonymous peer review

There are different levels of bi-directional anonymity throughout the peer review process, including whether or not the referees know who the authors are but not vice versa (single blind, the most common; (Ware, 2008)), or whether both parties remain anonymous to each other (double blind) (Table 1). Traditional double-blind review is based on the idea that peer evaluations should be impartial and based on the research, not ad hominem, but there has been considerable discussion over whether reviewer identities should remain anonymous (e.g., Snodgrass (2007); Baggs et al. (2008); Pontille and Torny (2014)) (Figure 3). Models such as triple-blind peer review even go a step further, where authors and their affiliations are reciprocally anonymous to the handling editor and the reviewers. This attempts to nullify the effects of one's scientific reputation, institution, or location on the peer review process, and is employed at the Open Access journal *Science Matters* (`sciencematters.io`) launched in early 2016.

There exist strong, but often conflicting, arguments and attitudes for both sides of the anonymity debate (see e.g., Prechelt et al. (2017)). In theory, anonymous reviewers are protected from potential backlashes for expressing themselves fully and, therefore, are more likely to be more honest in their assessments. Further, there is some evidence to suggest that double blind review can increase the acceptance rate of women-authored articles in the published literature (Darling, 2015). However, this kind of anonymity can be difficult to protect as there are ways in which identities can be revealed, albeit non-maliciously, such as through language and phrasing, prior knowledge of the research and a specific angle being taken, previous presentation at a conference, or even simple Web-based searches.

So while there is much potential value in anonymity, the corollary is also problematic in that anonymity can lead to reviewers being more aggressive, biased, negligent, orthodox, entitled, and politicized in their language and evaluation as they have no fear of negative consequences for their actions other than from the editor. (Weicher, 2008; Lee et al., 2013). Furthermore, by protecting the referees' identities journals lose an aspect of the prestige, quality, and validation in the review process, leav-

ing researchers to guess or assume this important aspect post-publication. The transparency associated with signing peer reviews aims to avoid competition and conflicts that can potentially arise due to the fact that referees are often the closest competitors to the authors, as they will naturally tend to be the most competent to assess the research (Campanario, 1998a,b). Eponymous peer review has the potential to encourage increased civility, accountability, and more thoughtful reviews (Boldt, 2011; Lipworth et al., 2011; Cope and Kalantzis, 2009; Fitzpatrick, 2010; Mulligan et al., 2013; Janowicz and Hitzler, 2012), as well as extending the process to become more of an ongoing, community-driven dialogue rather than a singular, static event (Maharg and Duncan, 2007; Bornmann et al., 2012). However, at the same time there is scope for the peer review to become less critical, skewed, and, as a result, become biased by community selectivity. If the anonymity of the reviewers is removed while maintaining author anonymity at any time during peer review, a skew and extreme accountability is imposed upon the reviewers, while authors remain relatively protected from any potential prejudices against them. However, such transparency provides, in theory, a mode of validation and should mitigate corruption as any association between authors and reviewers would be exposed. Yet, this approach has a clear disadvantage, in that accountability becomes extremely one-sided. Another possible result of this is that reviewers could be stricter in their appraisals within an already conservative environment, and thereby further prevent the publication of research.

2.4.1 Reviewing the evidence

Baggs et al. (2008) investigated the beliefs and preferences of reviewers about blinding. Their results showed double-blinding was preferred by 94% of reviewers, although some identified advantages to an un-blinded process. When author names were blinded, 62% of reviewers could not identify the authors while 17% could identify authors ≤10% of the time. Walsh et al. (2000) conducted a survey in which 76% of reviewers agreed to sign their reviews. In this case, signed reviews were of higher quality, were more courteous, and took longer to complete than unsigned reviews. Reviewers who signed were also more likely to recommend publication. In their study to explore the review process from the reviewers' perspectives, Snell and Spencer (2005) found that reviewers would be willing to sign their reviews and feel that the process should be transparent. Yet, a similar study by Melero and Lopez-Santovena (2001) found that 75% of surveyed respondents were in favor of reviewer anonymity, while only 17% were against it.

A randomized trial showed that blinding reviewers to the identity of authors improved the quality of the reviews (McNutt et al., 1990). This trial was repeated on a larger scale by Justice et al. (1998) and Van Rooyen et al. (1999), with neither study finding that blinding reviewers improved the quality of reviews. These studies also showed that blinding is difficult in practice as

many manuscripts include clues on authorship. Jadad et al. (1996) analyzed the quality of reports of randomized clinical trials and concluded that blind assessments produced significantly lower and more consistent scores than open assessments. The majority of additional evidence suggests that anonymity has little impact on the quality or speed of the review or of acceptance rates (Justice et al., 1998; van Rooyen et al., 1998; Isenberg et al., 2009), but revealing the identity of reviewers may lower the likelihood that someone will accept an invitation to review (Van Rooyen et al., 1999). Revealing the identity of the reviewer to a co-reviewer also has a small, editorially insignificant, but statistically significant beneficial effect on the quality of the review (van Rooyen et al., 1998). Authors who are aware of the identity of their reviewers may also be less upset by hostile and discourteous comments (McNutt et al., 1990). Other research found that signed reviews were more polite in tone, of higher quality, and more likely to ultimately recommend acceptance (Walsh et al., 2000).

2.4.2 The dark side of identification

This debate of signed versus unsigned reviews is not to be taken lightly. Early career researchers in particular are some of the most conservative in this area as they may be afraid that by signing overly critical reviews (i.e., those which investigate the research more thoroughly), they will become targets for retaliatory backlashes from more senior researchers. In this case, the justification for a reviewer anonymity is to protect junior researchers, as well as other marginalized demographics, from bad behaviour. Furthermore, author anonymity could potentially save junior authors from public humiliation from more established members of the research community, should they make errors in their evaluations. These potential issues are at least a part of the cause towards a general attitude of conservatism from the research community towards OPR. Indeed, they come up as the most prominent resistance factor in almost every formal discussion on the top of open peer review (e.g., van Rooyen et al. (1998); McCormack (2009); Godlee et al. (1998); Darling (2015); Pontille and Torny (2014); Snodgrass (2007)). However, it is not immediately clear how this widely-exclaimed but poorly documented potential abuse of signed-reviews is any different from what would occur in a closed system anyway, as anonymity provides a potential mechanism for referee abuse. The fear that most backlashes would be external to the peer review itself, and indeed occur in private, is probably the main reason why such abuse has not been widely documented. However, it can also be argued that by reviewing with the prior knowledge of open identification, such backlashes are prevented since researchers do not want to tarnish their reputations in a public forum. Under these circumstances, openness becomes a means to hold both referees and authors accountable for their public discourse, as well as making the editors' decisions on referee and publishing choice public. Either way, there is little documented evidence

that such retaliations actually occur either commonly or systematically. If they did then publishers that employ this model, such as *Frontiers* or *BioMed Central*, would be under serious question, instead of thriving as they are.

In an ideal world, we would expect that strong, honest, and constructive feedback are well received by authors, no matter their career stage. Yet, it seems that this is not the case. At least, there is the very real perception that it is not, and this is just as important from a social perspective. Retaliations to referees in such a negative manner represent serious cases of academic misconduct (Fox, 1994; Rennie, 2003). However, it must be made clear that this is not a direct consequence of OPR, but instead a failure of the general academic system to mitigate and act against inappropriate behavior. Increased transparency can only aid preventing and tackling the potential issues of abuse and publication misconduct, something which is almost entirely absent within a closed system. COPE provides advice to editors and publishers on publication ethics, and on how to handle cases of research and publication misconduct, including during peer review. COPE could be used as the basis for developing formal mechanisms adapted to innovative models of peer review, including those outlined in this paper. Any new OPR ecosystem could also draw on the experience accumulated by Online Dispute Resolution (ODR) researchers and practitioners over the past 20 years. ODR can be defined as "the application of information and communications technology to the prevention, management, and resolution of disputes" (Katsh and Rule, 2015), and could be implemented to prevent, mitigate, and deal with any potential misconduct during peer review alongside COPE. Therefore, the perceived danger of author backlash is highly unlikely to be acceptable in the current academic system, and if it does occur, it can be dealt with through increased transparency. Furthermore, bias and retaliation exist even in a double-blind review process (Baggs et al., 2008; Tomkins et al., 2017; Snodgrass, 2007), which is generally considered to be more conservative or protective. Such widespread identification of bias highlights this as a more general issue within peer review and academia more broadly, and we should be careful not to attribute it to any particular mode or trait of peer review. This is particularly relevant for more specialized fields, where the pool of potential authors and reviewers is relatively small (Riggs, 1995). Nonetheless, careful engagement with researchers, especially high-risk or marginalized communities, should be a necessary and vital step prior to implementation of any system of reviewer transparency.

2.4.3 The impact of identification and anonymity on bias

One of the biggest criticisms levied at peer review is that, like many human endeavours, it is intrinsically biased and not the objective and impartial process many regard it to be. The question is no longer about whether or not it is biased, but to what extent it is in different social dimensions. One of the major issues is that peer review suffers from systemic confirmatory bias, with only results that are deemed as significant, statistically or otherwise, being selected for publication (Mahoney, 1977). This causes a distinct bias within the published research record (van Assen et al., 2014), as a consequence of perverting the research process itself by creating an incentive system that is almost entirely publication-oriented. Others have described the issues with such an asymmetric evaluation criteria as lacking the core values of a scientific process (Bon et al., 2017).

The evidence on whether there is bias in peer review against certain author demographics is mixed, but overwhelmingly in favor of systemic bias against women in article publishing (Budden et al., 2008; Helmer et al., 2017; Grivell, 2006; McKiernan, 2003; Tregenza, 2002; Lloyd, 1990; Lerback and Hanson, 2017; Smith, 2006; Darling, 2015; Roberts and Verhoef, 2016) (although see Blank (1991); Webb et al. (2008); Whittaker (2008)). After the journal *Behavioural Ecology* adopted double-blind peer review in 2001, there was a significant increase in accepted manuscripts by women first authors; an effect not observed in similar journals that did not change their peer review policy (Budden et al., 2008). One of the most recent public examples of this bias is the case where a reviewer told the authors that they should add more male authors to their study (Bernstein, 2015). More recently, it has been shown in the *Frontiers* journal series that women are under-represented in peer-review and that editors of both genders operate with substantial same-gender preference (Helmer et al., 2017). The most famous piece of evidence on bias against authors comes from a study by Peters and Ceci (1982) using psychology journals. They took 12 studies that came from prestigious institutions that had already been published in psychology journals. They retyped the papers, made minor changes to the titles, abstracts, and introductions but changed the authors' names and institutions. The papers were then resubmitted to the journals that had first published them. In only three cases did the journals realize that they had already published the paper, and eight of the remaining nine were rejected—not because of lack of originality but because of the perception of poor quality. Peters and Ceci (1982) concluded that this was evidence of bias against authors from less prestigious institutions, although the deeper causes of this bias remain unclear at the present. A similar effect was found in an orthopaedic journal by Okike et al. (2016), where reviewers were more likely to recommend acceptance when the authors' names and institutions were visible than when they were redacted. Further studies have shown that peer review is substantially positively biased towards authors from top institutions (Ross et al., 2006; Tomkins et al., 2017), due to the perception of prestige of those institutions and, consequently, of the authors as well. Further biases based on nationality and language have also been shown to exist (Ross et al., 2006; Dall'Aglio, 2006; Link, 1998; Ernst and Kienbacher, 1991; Tregenza, 2002).

While there are relatively few large-scale investigations of

the extent and mode of bias within peer review (although see Lee et al. (2013) for an excellent overview of the different levels in which bias can be potentially injected into the process), these studies together indicate that inherent biases are systemically embedded within the process, and must be accounted for prior to any further developments in peer review. This range of population-level investigations into attitudes and applications of anonymity, and the extent of any biases resulting from this, exposes a highly complex picture, and there is little consensus on its impact at a system-wide scale. However, based on these often polarised studies, it is inescapable to conclude that peer review is highly subjective, rarely impartial, and definitely not as homogeneous as it is often regarded.

Applying a single, blanket policy regarding anonymity would greatly degrade the ability of science to move forward, especially without a wide flexible to manage exceptions. The reasons to avoid one definite policy are the inherent complexity of peer review systems, the interplay with different cultural aspects within the various sub-sectors of research, and the difficulty in identifying whether anonymous or identified works are objectively better. As a general overview of the current peer review ecosystem, Nobarany and Booth (2016) recently recommended that, due to this inherent diversity, peer review policies and support systems should remain flexible and customizable to suit the needs of different research communities. We expect that, by emphasizing the different shared values across research communities, as well as their commonalities, we will see a new diversity of OPR processes developed across disciplines in the future. Remaining ignorant of this diversity of practices and inherent biases in peer review, as both social and physical processes, would be an unwise approach for future innovations.

2.5 Decoupling peer review from publishing

One proposal to transform scholarly publishing is to decouple the concept of the journal and its functions (e.g., archiving, registration and dissemination) from peer review and the certification that this provides. Some even hail this decoupling process as the "paradigm shift" that scholarly publishing needs (Priem and Hemminger, 2012). Some publishers, journals, and platforms are now taking a more adventurous exploration of peer review that occurs subsequently to publication (Figure 3). Here, the principle is that all research deserves the opportunity to be published (usually pending some form of initial editorial selectivity), and that filtering through peer review occurs subsequently to the actual communication of research articles (i.e., a publish then filter process). This is often termed "post-publication peer review", a confusing terminology based on what constitutes "publication" in the digital age, depending on whether it occurs on manuscripts that have ben previously peer reviewed or not (`blogs.openaire.eu/?p=1205`). Numerous venues now provide inbuilt systems for post-publication peer review, including *RIO, PubPub, ScienceOpen, The*

Winnower, and *F1000 Research*. In addition to the systems adopted by journals, other post-publication annotation and commenting services exist independent of any specific journal or publisher and operating across platforms, such as *hypothes.is*, *PaperHive*, and *PubPeer*.

Initiatives such as the *Peerage of Science* (`peerageofscience.org`), *RUBRIQ* (`rubriq.com`), and *Axios Review* (`axiosreview.org`; closed in 2017) have implemented a decoupled model of peer review. These tools work based on the same core principles as traditional peer review, but authors submit their manuscripts to the platforms first instead of journals. The platforms provide the referees, either via subject-specific editors or via self-managed agreements. After the referees have provided their comments and the manuscript has been improved, the platform forwards the manuscript and the referee reports to a journal. Some journal policies accept the platform reviews as if the reviews were coming from the journal's pool of reviewers, while others still require the journal's handling editor to look for additional reviewers. While these systems usually cost money for authors, these costs can sometimes be deducted from any publication fees once the article has been published. Journals accept deduction of these costs because they benefit by receiving manuscripts that have already been assessed for journal fit and have been through a round of revisions, thereby reducing their workload. A consortium of publishers and commercial vendors recently established the Manuscript Exchange Common Approach (MECA; `manuscriptexchange.org`) as a form of portable review in order to cut down inefficiency and redundancy. Yet, it still is in too early a stage to comment on its viability.

LIBRE (`openscholar.org.uk/libre`) is a free, multidisciplinary, digital article repository for formal publication and community-based evaluation. Reviewers' assessments, citation indices, community ratings, and usage statistics, are used by *LIBRE* to calculate multiparametric performance metrics. At any time, authors can upload an improved version of their article or decide to send it to an academic journal. Launched in 2013, *LIBRE* was subsequently combined with the *Self-Journal of Science* (`sjscience.org`) under the combined heading of *Open Scholar* (`openscholar.org.uk`). One of the tools that *Open Scholar* offers is a peer review module for integration with institutional repositories, which is designed to bring research evaluation back into the hands of research communities themselves (`openscholar.org.uk/open-peer-review-module-for-repositories/`). *Academic Karma* is another new service that facilitates peer review of pre-prints from a range of sources (`academickarma.org/`.)

2.5.1 Pre-prints and overlay journals

In fields such as mathematics, astrophysics, or cosmology, research communities already commonly publish their work on *arXiv* (Larivière et al., 2014). To date, this
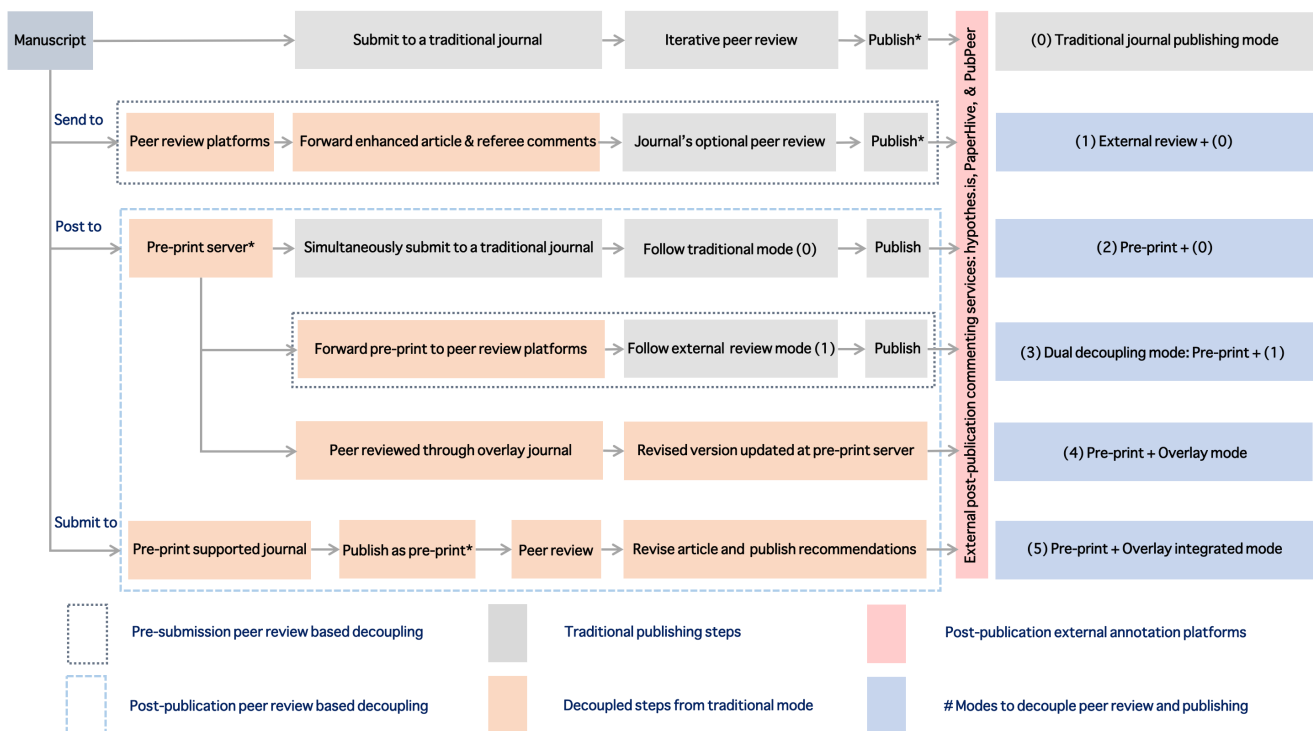
Figure 3. **Traditional versus different decoupled peer review models: Peer review under decoupled model either happens pre-submission or post-publication. The dotted border lines in the figure highlight this element, with boxes colored in orange representing decoupled steps from the traditional publishing model and the ones colored gray depicting the traditional publishing model itself. Pre-submission peer review based decoupling (1) offers a route to enhance a manuscript before submitting it to a traditional journal; pre-print first mode of decoupling (2, 3, 4, and 5) offer four different ways to post-publication peer review for revision and acceptance. Dual-decoupling (3) is when a manuscript first posted as a pre-print (primary decoupling) is sent for external peer review (secondary decoupling) before its initial submission to a traditional journal. The asterisks in the figure indicate when the manuscript first enters the public view irrespective of its peer review status.**

platform has accumulated more than one million research documents – pre-prints or e-prints – and currently receives 8000 submissions a month with no costs to authors. *arXiv* also sparked innovation for a number of communication and validation tools within restricted communities, although these seem to be largely local, non-interoperable, and do not appear to have disrupted the traditional scholarly publishing process to any great extent (Marra, 2017). In other fields, the uptake of pre-prints has been relatively slower, although it is gaining momentum with the development of platforms such as *bioRxiv* and several newly established ones through the *Center for Open Science*, including *engrXiv* (engrXiv.org) and *psyarXiv* (psyarxiv.com), and social movements such as ASAPBio (asapbio.org). Manuscripts submitted to these pre-print servers are typically a draft version prior to formal submission to a journal for peer review. Primary motivation for this is the lengthy time taken for peer review and formal publication, and causes the timing of peer review to occur subsequent to making manuscripts public. However, sometimes these articles are not submitted anywhere else and form what some regard as grey literature (Luzi, 2000). Papers on digital repositories are cited on a daily basis and much research builds upon them,

although they may suffer from a stigma of not having the scientific stamp of approval of peer review (Adam, 2010). Some journal policies explicitly attempt to limit their citation in peer-reviewed publications (e.g., *Nature* nature.com/nature/authors/gta/#a5.4), *Cell* cell.com/cell/authors), and recently the scholarly publishing sector even attempted to discredit their recognition as valuable publications (asapbio.org/faseb). In spite of this, the popularity and success of pre-prints is testified by their citation records with four of the top five venues in physics and maths being *arXiv* sub-sections (scholar.google.com/citations?view_op=top_venues&hl=en&vq=phy). Similarly, the single most highly cited venue in economics being the *NBER Working Papers* server (scholar.google.com/citations?view_op=top_venues&hl=en&vq=bus_economics), according to the *Google Scholar* h5-index.

The overlay journal, first described by Ginsparg (1997), is a novel type of journal that operates by having peer review as an additional layer on top of pre-prints. These have built on the concept of deconstructed journals (Smith, 1999), which decouple peer-review from publishing (Hettyey et al., 2012; Stemmle and Collier, 2013; Vines, 2015b; Patel, 2014). New overlay journals such

as *The Open Journal* (`theoj.org`) or *Discrete Analysis* (`discreteanalysisjournal.com`) are exclusively peer review platforms that circumvent traditional publishing by utilizing the pre-existing infrastructure and content of pre-print servers like *arXiv*. Peer review here is performed easily, rapidly, and cheaply subsequent to the initial publication of articles. The reason they are termed "overlay" journals is that the articles remain on *arXiv* in their peer reviewed state, with the "journals" mostly comprising a simple list of links to these versions (Gibney, 2016).

A similar approach to that of overlay journals is being developed by *PubPub* (pubpub.org), which allows authors to self-publish their work. *PubPub* then provides a mechanism for creating overlay journals that can draw from and curate the content hosted on the platform itself. This model incorporates the pre-print server and final article publishing into one contained system. *EPISCIENCES* is another platform that facilitates the creation of peer reviewed journals with their content hosted on digital repositories (Berthaud et al., 2014). *ScienceOpen* provides editorially-managed collections of articles drawn from pre-prints and a combination of open access and non-open venues (e.g., scienceopen.com/collection/Science20). Editors compile articles to form a collection, write an editorial, and can invite referees to peer review the articles. This process is mediated by *ORCID* for quality control, and *CrossRef* and *Creative Commons* licensing for appropriate recognition. They are essentially equivalent to community-mediated overlay journals, but with the difference that they also draw on additional sources beyond pre-prints.

2.5.2 Two-stage peer review and Registered Reports

Registered Reports represent a significant departure from conventional peer review in terms of relative timing and increased rigour (Chambers et al., 2014, 2017; Nosek and Lakens, 2014). Here, peer review is split into two stages. Research questions and methodology (i.e., the study design itself) are subject to a first round of evaluation prior to any data collection or analysis taking place (Figure 4). If a protocol is found to be of sufficient quality to pass this stage, the study is then provisionally accepted for publication. Once the research has been completed and written-up, completed manuscripts are then subject to a second-stage of peer review which, in addition to affirming the soundness of the results, also confirms that data-collection and analysis occurred in accordance with the originally described methodology. The format, originally introduced by the psychology journals *Cortex* and *Perspectives in Psychological Science* in 2013, is now used in some form at more than 40 journals (Nature Human Behaviour, 2017). Registered Reports are designed to boost research integrity by ensuring the publication of all research results and, hence, aids the elimination of publication bias. As opposed to the traditional model of publication, where "positive" results are more likely to be published, results

remain unknown at the time of review and therefore even "negative" results are equally as likely to be published. Such a process is designed to incentivize data-sharing, guards against dubious practices such as selective reporting of results (via so-called "p-hacking" and "HARKing"—Hypothesizing After the Results are Known) and low statistical power, and also prioritizes accurate reporting over that which is perceived to be of higher impact or publisher worthiness.

2.5.3 Peer Review by Endorsement

A relatively new mode of named pre-publication review is that of pre-arranged and invited review, originally proposed as author-guided peer review (Perakakis et al., 2010), which *ScienceOpen* terms Peer Review by Endorsement (PRE) (`about.scienceopen.com/peer-review-by-endorsement-pre/`). This has also been implemented at *RIO*, and is functionally similar to the Contributed Submissions of *PNAS* (`pnas.org/site/authors/editorialpolicies.xhtml#contributed`). This model requires an author to solicits reviews from their peers prior to submission in order to assess the suitability of a manuscript for publication. While some might see this as a potential bias, it is worth bearing in mind that many journals already ask authors who they want to review their papers, or who they should exclude. To avoid potential pre-submission bias, reviewer identities and their endorsements are made publicly available alongside manuscripts, which also removes any possible deleterious editorial criteria from inhibiting the publication of research. Also, PRE is much cheaper, legitimate, unbiased, faster, and more efficient alternative to the traditional publisher-mediated method. In theory, depending on the state of the manuscript, this means that submissions can be published much more rapidly, as relatively little processing is required. PRE also has the potential advantage of being more useful to non-native English speaking authors by allowing them to work with editors and reviewers in their first languages.

Endorsements and recommendations are a form of peer review that can facilitate re-use of published works. This has been most evident in the Open Educational Resources (OER) movement, in which peer review and testimonials on Open Education repositories, such as *Merlot*, form a way to filter the many resources available. Peer review, including recommendations, has been effectively utilized in the creation and sharing of Open Textbooks. Petrides et al. (2011) and Harley et al. (2010) found that proof of peer review by trusted experts was a significant factor leading to adoption of textbooks by instructors who express concern about the quality of a free textbook. Some OER reviewers are even paid for their reviews (Open Access Textbook Task Force, 2010), while other reviews are done by volunteer editors and the users of the resources. (`info.merlot.org/merlothelp/merlot_peer_review_information.htm`).
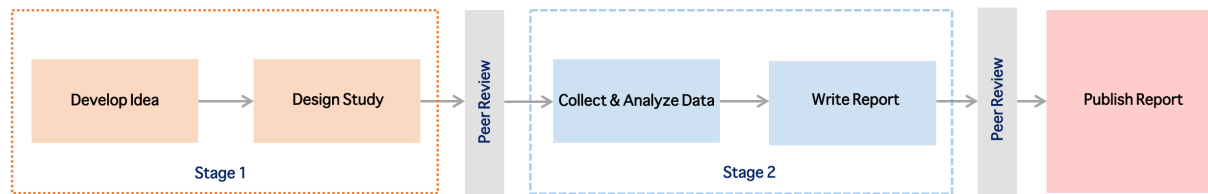
Figure 4. The process of Registered Reports. Each peer review stage also includes editorial input.

2.5.4 Limitations of decoupled peer review

Despite a general appeal for post-publication peer review and considerable innovation in this field, the appetite among researchers is limited reflecting an overall lack of engagement with the process (e.g., Nature (2010)). As recently as 2012, it was reported that relatively few platforms allow users to evaluate manuscripts post-publication (Yarkoni, 2012). Even platforms such as *PLOS* have a restricted scope and limited user base: analysis of publicly available usage statistics indicate that at the time of writing, *PLOS* articles have each received an average of 0.06 ratings and 0.15 comments (see also Ware (2011)). Part of this may be due to how post-publication peer review is perceived culturally, with the name itself being anathema and considered an oxymoron, as most researchers usually consider a published article to be one that has already undergone formal peer review. At the present, it is clear that while there are numerous platforms providing decoupled peer review services, these are largely non-interoperable. The result of this, especially for post-publication services, is that most evaluations are difficult to discover, lost, or rarely available in an appropriate context or platform for re-use. To date, it seems little effort has been focused on aggregating the content of these services, which hinders its recognition as a valuable community process and for additional evaluation or assessment decisions.

While several new overlay journals are currently thriving, the history of their success is invariably limited, and most journals that experimented with the model returned to their traditional coupled roots (Priem and Hemminger, 2012). *Axios Review* was closed down in early 2017 due to a lack of uptake from researchers, with the founder stating: "I blame the lack of uptake on a deep inertia in the researcher community in adopting new workflows" (Davis, 2017). Finally, it is probably worth mentioning that not a single overlay journal appears to have emerged outside of physics and math (Priem and Hemminger, 2012). This is despite the fast growth of *arXiv* spin-offs like *biorXiv*, and potential layered peer review through services such as *ScienceOpen* or the recently launched *Peer Community In* (peercommunityin.org).

Coupled with the demise of services such as *Axios Review*, the generally low uptake of decoupled peer review processes suggests the overall reluctance of many research communities to adapt outside of the traditional coupled model. In this section, we have discussed a range of different arguments, variably successful platforms, and surveys and reports about peer review. Taken together, these reveal an incredible amount of friction to experimenting with peer review beyond that which is typically and incorrectly viewed as the only way of doing it. This reluctance is emphasized in recent surveys, such as that by Ross-Hellauer (2017), which suggest that while attitudes towards the principles of OPR are rapidly becoming more positive, faith in its execution is not. We can perhaps expect this divergence due to the rapid pace of innovation, which has not led to rigorous or longitudinal evidence that these models are superior to the traditional process at either a population or system-wide level. Cultural or social inertia, then, is defined by this cycle between low uptake and limited evidence of optimal alternative models, irrespective of the flaws associated with the present system. Perhaps more important is the general underappreciation of this intimate relationship between social and technological barriers, that is undoubtedly required to overcome this cycle. The proliferation of social media over the last decade provide excellent examples to peer review for how digital communities can leverage new technologies for great effect.

3 Potential future models

As we have discussed in detail above, there has been considerable technological innovation in peer review in the last decade, which is leading to critical examination of it as a social process. Much of this has been driven by the advent of Web 2.0 technologies and new social media platforms, and an overall shift towards a more open system of scholarly communication. Previous work in this arena has described features of a *Reddit*-like model combined with additional personalized features of other social platforms, like *Stack Exchange*, *Netflix*, and *Amazon* (Yarkoni, 2012). Here, we develop upon this by considering additional traits of models such as *Wikipedia*, *GitHub*, *Blockchain*, and discuss these in the rapidly evolving socio-technological environment for the present system of peer review. In any vision of the future of scholarly publishing (Kriegeskorte et al., 2012), the evolution of peer review and evaluation systems must be considered. Any future peer review platform or system would greatly benefit from considering the following key features:

1. Quality control and moderation, possibly through openness and transparency;

2. Certification via personalized reputation or performance metrics;

3. Incentive structures to motivate and encourage engagement.

While discussing a number of principles that should guide the implementation of novel platforms for evaluating scientific work, Yarkoni (2012) argued that many of the problems researchers face have already been successfully addressed by a range of non-research focused social Web applications. Therefore, developing next-generation platforms for scientific evaluations should focus on adapting the best currently used approaches for these rather than on innovating entirely new ones (Neylon and Wu, 2009; Priem and Hemminger, 2010; Yarkoni, 2012). One important element that will determine the success or failure of any such peer-to-peer reputation or evaluation system is a critical mass of researcher uptake. This has to be carefully balanced with the demands and uptakes of restricted scholarly communities, which have inherently different motivations and practices in peer review. A remaining issue is the aforementioned cultural inertia, which can lead to low adoption of anything innovative or disruptive to traditional workflows in research. This is a perfectly natural trait for communities, where ideas out-pace technological innovation, which in turn out-paces the development of social norms. Hence, rather than proposing an entirely new platform or model of peer review, our approach here is to consider the advantages and disadvantages of existing models and innovations in social services and technologies (Table 4). We use this to explore the ways in which such traits can be adapted, combined, and applied to build a more effective and efficient peer review system, while potentially reducing friction to its uptake.

3.1 A Reddit-based model

Reddit (reddit.com) is an open-source, community-based platform where users submit comments and original or linked content, organized into thematic lists of subreddits. As Yarkoni (2012) noted, a thematic list of subreddits can be automatically generated for any peer review platform using keyword metadata generated from sources like the National Library of Medicine's Medical Subject Heading (MeSH) ontology. Members, or redditors, can upvote or downvote any submissions based on quality and relevance, and publicly comment on all shared content. Individuals can subscribe to contribution lists, and articles can be organized by time (newest to oldest) or level of engagement. Quality control is invoked by moderation through subreddit mods, who can filter and remove inappropriate comments and links. A score is given for each link and comment as the sum of upvotes minus downvotes, thus providing an overall ranking system. At *Reddit*, highly scoring submissions are relatively ephemeral, with an automatic down-voting algorithm implemented that shifts them further down lists as new content is added, typically within 24 hours of initial posting.

3.1.1 Reddit as an existing "journal" of science

The subreddit for Science (reddit.com/r/science) is a highly-moderated discussion channel, curated by at least 600 professional researchers and with more than 15 million subscribers at the time of writing. The forum has even been described as "The world's largest 2-way dialogue between scientists and the public" (Owens, 2014). Contributors here can add flair to their posts as a way of thematically organizing them based on research discipline, analogous to the container function of a typical journal. Individuals can also have flair as a form of subject-specific credibility (i.e., a peer status) upon provision of proof of education in their topic. Public contributions from peers are subsequently stamped with a status and area of expertise, such as "Grad student|Earth Sciences."

Scientists already further engage with *Reddit* through science AMAs (Ask Me Anythings), which tend to be quite popular. However, the level of discourse provided in this is generally not equivalent in depth compared to that perceived for peer review, and is more akin to a form of science communication or public engagement with research. In this way, *Reddit* has the potential to drive enormous amounts of traffic to primary research and there even is a phenomenon known as the "*Reddit* hug of death", whereby servers become overloaded and crash due to *Reddit*-based traffic. The */r/science* subreddit is viewed as a venue for "scientists and lay audiences to openly discuss scientific ideas in a civilized and educational manner", according to the organizer, Dr. Nathan Allen (Lee, 2015). As such, an additional appeal of this model is that it could increase the public level of scientific literacy and understanding.

3.1.2 Reddit-style peer evaluation

The essential part of any *Reddit*-style model with potential parallels to peer review is that links to scientific research can be shared and ranked (upvoted or downvoted) by the community. All links or texts can be publicly discussed in terms of methods, context, and implications, similar to any post-publication commenting system. Such a process for peer review could essentially operate as an additional layer on top of a pre-print archive or repository, much like a social version of an overlay journal. Ultimately, a public commenting system like this could achieve the same depth of peer evaluation as the formal process, but as a crowd-sourced process. However, it is important to note here that this is a mode of instantaneous publication prior to peer review, with filtering through interaction occurring post-publication. Furthermore, comments can receive similar treatment to submitted content, in that they can be upvoted, downvoted, and further commented upon in a cascading process. An advantage of this is that multiple comment threads can form on single posts and viewers can track individual discussions. Here, the highest-ranked comments could simply be presented at the top of the thread, while those of lowest ranking

Table 4. Potential pros and cons of the main features of the investigated models. Note that some of these are already employed by different research platforms.

Feature	Description	Pros	Cons/Risks	Existing models
Voting or rating	Quantified review evaluation (5 stars, points), including up- and down-votes	Community-driven, quality filter, simple and efficient	Randomized procedure, auto-promotion, gaming, popularity bias, non-static	Reddit, Stack Exchange, Amazon
Openness	Public visibility of review content	Responsibility, accountability, context, higher quality	Peer pressure, potential lower quality, invites retaliation	All
Reputation	Reviewer evaluation and ranking (points, review statistics)	Quality filter, reward, motivation	Imbalance based on user status, encourages gaming, platform-specific	Stack Exchange, GitHub, Amazon
Public commenting	Visible comments on paper/review	Living/organic paper, community involvement, progressive, inclusive	Prone to harassment, time consuming, non-interoperable, low re-use	Reddit, Stack Exchange, Hypothesis
Version control	Managed releases and configurations	Living/organic objects, verifiable, progressive, well-organized	Citation tracking, time consuming, low trust of content	GitHub, Wikipedia
Incentivization	Encouragement to engage with platform and process via badges/money or recognition	Motivation, return on investment	Research monetization, can be perverted by greed, expensive	Stack Exchange, Blockchain
Authentication and certification	Filtering of contributors via verification process	Fraud control, author protection, stability	Hacking, difficult to manage	Blockchain
Moderation	Filtering of inappropriate behavior in comments, rating	Community-driven, quality filter	Censorship, mainstream speech	Reddit, Stack Exchange

remain at the bottom.

In theory, a subreddit could be created for any sub-topic within research, and a simple nested hierarchical taxonomy could make this as precise or broad as warranted by individual communities. *Reddit* allows any user to create their own subreddit, pending certain status achievements through platform engagement. In addition, this could be moderated externally through *ORCID*, similar to the approach taken by *ScienceOpen*, in which five items in a peer's ORCID profile are required to perform a peer review; or in this case, create a new subreddit. Connection to a social network within academia, such as *ORCID*, further allows community validation, verification, and judgement of importance. For example, being able to see whether senior figures in a given field have read or upvoted certain threads can be highly influential in decisions to engage with that thread, and vice versa. A very similar process already occurs at the *Self Journal of Science*, where contributors have a choice of voting either "This article has reached scientific standards" or "This article still needs revisions", with public disclosure of who has voted in either direction. Threaded commenting could also be implemented as it is vital to the success of any collaborative filtering platform, and also provides a highly efficient corrective mechanism. Peer evaluation in this form emphasizes progress and research as a discourse over piecemeal publications or objects as part of a lengthier process. Such a system could be applied to other forms of scientific work, which includes code, data and images, thereby allowing contributors to claim credit for their full range of research outputs. Comments could be signed by default, pseudonymous, or anonymized until a contributor

chooses to reveal their identity. If required, anonymized comments could be filtered out automatically by users. A key to this could be peer identity verification, which can be done at the back-end via email or integrated via *ORCID*.

3.1.3 Translating engagement into prestige
Reddit karma points are awarded for sharing links and comments, and having these upvoted or downvoted by other registered members. The simplest implementation of such a voting system for peer review would be through interaction with any article in the database with a single click. This form of field-specific social recommendation for content simultaneously creates both a filter and a structured feed, similar to *Facebook* and *Google+*, and can easily be automated. With this, contributions get a rating, which accumulate to form a peer-based rating as a form of reputation and translate into a quantified level of community-granted prestige. Ratings are transparent and contributions and their ratings can be viewed on a public profile page. More sophisticated approaches could include graded ratings—e.g., five-point responses, like those used by *Amazon*—or separate rating dimensions providing peers with an immediate snapshot of the strengths and weaknesses of each article. Such a system is already in place at *ScienceOpen*, where referees evaluate an article for importance, validity, completeness, and comprehensibility using a five-star system. For any given set of articles retrieved from the database a ranking algorithm could be used to dynamically order articles on the basis of a combination of quality (an article's aggregate rating within the system, like at *Stack Exchange*), relevance (using a recommendation system akin to *Amazon* or *ScienceOpen*), and recency (newly added articles could

receive a boost). By default, the same algorithm would be implemented for all peers, as on *Reddit*. The issue here is making any such karma points equivalent to the amount of effort required to obtain them, and also valued by the broader research community and assessment bodies. This could be facilitated through a simple badge incentive system, such as that designed by the *Center for Open Science* for core open practices (`cos.io/our-services/open-science-badges/`).

3.1.4 Can the wisdom of crowds work with peer review?

One might consider a *Reddit*-style model as pitching quantity versus quality. Typically, comments provided on *Reddit* are not at the same level in terms of depth and rigor as those that we would expect from traditional peer review—as in, there is more to research evaluation than simply upvoting or downvoting. Furthermore, the range of expertise is highly variable due to the inclusion of specialists and non-specialists as equals ("peers") within a single thread. However, there is no reason why a user prestige system akin to *Reddit* flair cannot be utilised to differentiate varying levels of expertise. The primary advantage here is that the number of participants is uncapped, therefore emphasizing the potential that *Reddit* has in scaling up participation in peer review. With a *Reddit* model, we must hold faith that sheer numbers will be sufficient in providing an optimal assessment of any given contribution and that any such assessment will ultimately provide a consensus of high quality and reusable results. Social review of this sort must therefore consider at what point is the process of review constrained in order to produce such a consensus, and one that is not self-selective as a factor of engagement rather than accuracy. This is termed the "Principle of Multiple Magnifications" by Kelty et al. (2008), which surmises that in spite of self-selectivity, more reviewers and more data about them will always be better than fewer reviewers and less data. The additional challenge, then, will be to capture and archive consensus points for external re-use. Journals such as *F1000 Research* already have such a tagging system, where reviewers can mark a submission as approved after peer review iterations.

One potential phenomenon for this style of system is that of "the rich get richer". Here, content from more prominent researchers will receive relatively more comments and ratings, and ultimately hype, as with any hierarchical system including that for traditional scholarly publishing. The opposite might also be true in that research from unknown authors will go relatively under-noticed and under-used, but will at least have been publicized. One solution to this draws on the science subreddit, by having a core community of editors. These could be empowered to invite peers to contribute to discussion threads, essentially wielding the same executive power as a journal editor, but combined with that of a forum moderator. Recent evidence suggests that such intelligent crowd reviewing has the potential to be an efficient and high quality pro-

cess (List, 2017).

3.2 An Amazon-style rate and review model

Amazon was one of the first websites allowing the posting of public, customer book reviews. The process is completely open and informal in that anyone can write a review and vote, providing usually that they have purchased the product. Customer reviews of this sort are peer-generated product evaluations hosted on a third-party website, such as *Amazon* (Mudambi and Schuff, 2010). Here, usernames can be either real identities or pseudonyms. Reviews can also include images, and have a header summary. In addition, a fully searchable question and answer section on individual product pages allows users to ask specific questions, answered by the page creator, and voted on by the community. Top-voted answers are then displayed at the top. Chevalier and Mayzlin (2006) investigated the *Amazon* review system finding that, while reviews on the site tended to be more positive, negative reviews had a greater impact in determining sales. Reviews of this sort can therefore be thought of in terms of value addition or subtraction to a product or content, and ultimately used to guide a third-party evaluation of a product and purchase decision (i.e., a selectivity process).

3.2.1 Amazon's star-rating system

Star-rating systems are used frequently at a high-level in academia, and commonly used to define research excellence, albeit perhaps in a flawed and an arguably detrimental way; e.g., the Research Excellence Framework in the UK (`ref.ac.uk`) (Murphy and Sage, 2014; Mhurchú et al., 2017; Moore et al., 2017). A study about Web 2.0 services and their use in alternative forms of scholarly communication by UK researchers found that nearly half (47%) of those surveyed expected that peer review would be complemented by citation and usage metrics and user ratings in the future (Procter et al., 2010b,a). *Amazon* provides a sophisticated collaborative filtering system based on five-star ratings, usually combined with several lines of comments and timestamps. This system is summarized with the proportion of total customer reviews that have rated a product at each star level. An average star rating is also given for each piece of content. A low rating (one star) indicates an extremely negative view, whereas a high rating (five stars) reflects a positive view of the product. An intermediate scoring (three stars) can either represent a mid-view of a balance between negative and positive points, or merely reflect a nonchalant attitude towards a product. These ratings reveal fundamental details of accountability and are a sign of popularity and quality for items and sellers.

However, the utility of such a star-rating system for research is not immediately clear, or whether positive, moderate, or negative ratings would be more useful. A rating by itself would be a fairly useless design for researchers without being able to see the context and justification behind it. It is also unclear how a combined rate and review

system would work for non-traditional research outputs, as the extremity and depth of reviews have been shown to vary depending on the type of content (Mudambi and Schuff, 2010). Furthermore, the ubiquitous five-star rating tool used across the Web is flawed in practice and produces highly skewed results. For one, when people rank products or write reviews online, they are more likely to leave positive feedback. The vast majority of ratings on *YouTube*, for instance, is five stars and it turns out that this is repeated across the Web with an overall average estimated at about 4.3 stars, no matter the object being rated (Crotty, 2009). Ware (2011) confirmed this average for articles rated in *PLOS*, suggesting that academic ranking systems operate in a similar manner to other social platforms. Ratings systems also select for popularity rather than quality, which is the opposite of what scholarly evaluation seeks (Ware, 2011). Another problem with commenting and rating systems is that they are open to gaming and manipulation. The *Amazon* system has been widely abused and it has been demonstrated how easy it is for an individual or small groups of friends to influence the popularity metrics even on hugely-visited websites like Time 100 (Harmon and Metaxas, 2010; Emilsson, 2015). *Amazon* has historically prohibited compensation for reviews, prosecuting businesses who pay for fake reviews as well as the individuals who write them. Yet, with the exception that reviewers could post an honest review in exchange for a free or discounted product as long as they disclosed that fact. A recent study of over seven million reviews indicated that the average rating for products with these incentivized reviews was higher than non-incentivized ones (Review Meta, 2016). Aiming to contain this phenomenon, *Amazon* has recently decided to adapt its Community Guidelines to eliminate incentivized reviews. As mentioned above, *ScienceOpen* offers a five-star rating system for papers, combined with post-publication peer review, but here the incentive is simply that the review content can be re-used, credited, and cited. How this translates to user and community perception in an academic environment remains an interesting question for further research.

3.2.2 Reviewing the reviewers

At *Amazon*, users can vote whether or not a review was helpful with simple binary yes or no options. Potential abuse can also be reported and avoided here by creating a system of community-governed moderation. After a sufficient number of yes votes, a user is upgraded to a spotlight reviewer through what essentially is a popularity contest. As a result, their reviews are given more prominence. Top reviews are those, which receive the most helpful upvotes, usually because they provide more detailed information about a product.

One potential way of improving rating and commenting systems is to weight such ratings according to the reputation of the rater (as done on *Amazon*, *eBay*, and *Wikipedia*). Reputation systems are intended to achieve three things: to foster good behavior, penalize bad behav-

ior, and reduce the risk of harm to others as a result of bad behavior (Ubois, 2003). Key features are that reputation can rise and fall and that reputation is based on behavior rather than social connections, thus prioritizing engagement over popularity. In addition, reputation systems do not have to use the true names of the participants but, to be effective and robust, they must be tied to an enduring identity infrastructure. Frishauf (2009) proposed a reputation system for peer review in which the review would be undertaken by people of known reputation, thereby setting a quality threshold that could be integrated into any social review platform and automated (e.g., via *ORCID*). One further problem with reputation systems is that having a single formula to derive reputation leaves the system open to gaming, as with almost any process that can be measured and quantified. Gashler (2008) proposed a decentralized and secured system where each reviewer would digitally sign each paper, hence the digital signature would link the review with the paper. Such a web of reviewers and papers could be data mined to reveal information on the influence and connectedness of individual researchers within the research community. Depending on how the data were mined, this could be used as a reputation system or web-of-trust system that would be resistant to gaming because it would specify no particular metric.

3.3 A Stack Exchange/Overflow-style model

Stack Exchange (`stackexchange.com`) is a collective intelligence system comprising multiple individual question and answer sites, many of which are already geared towards particular research communities, including maths and physics. The most popular site within *Stack Exchange* is *Stack Overflow*, a community of software developers and a place where professionals exchange problems, ideas, and solutions. *Stack Exchange* works by having users publish a specific problem, and then others contribute to a discussion on that issue. This format is considered to be a form of dynamic publishing by some (Heller et al., 2014). The appeal of *Stack Exchange* is that threaded discussions are often brief, concise, and geared towards solutions, all in a typical Web forum format. Highly regarded answers are positioned towards the top of threads, with others concatenated beneath. Like the *Amazon* model of weighted ratings, voting in *Stack Exchange* is more of a process that controls relative visibility. The result is a library of topical questions with high quality discussion threads and answers, developed by capturing the long tail of knowledge from communities of experts. The main distinction between this and scholarly publishing is that new material rarely is the focus of discussion threads. However, the ultimate goal remains the same: to improve knowledge and understanding of a particular issue. As such, *Stack Exchange* is about creating self-governing communities and a public, collaborative knowledge exchange forum based on software (Begel et al., 2013).

3.3.1 Existing Overflow-style platforms

Some subject-specific platforms for research communities already exist that are similar to or based on *Stack Exchange* technology. These include *BioStars* (`biostars.org`), a rapidly growing Bioinformatics resource, the use of which has contributed to the completion of traditional peer reviewed publications (Parnell et al., 2011). Another is *PhysicsOverflow*, a platform for real-time discussions between physics professionals combined with an open peer review system (Pallavi Sudhir and Knöpfel, 2015). *PhysicsOverflow* forms the counterpart forum to *MathOverflow* (Tausczik et al., 2014), with both containing a graduate-level question and answer forum, and an Open Problems section for collaboration on research issues. Both have a Reviews section to complement formal journal-led peer review, where peers can submit pre-prints (e.g., from *arXiv*) for public peer evaluation, and is considered to be an "*arXiv*-2.0". Responses are divided into reviews and comments, and given a score based on votes for originality and accuracy. Similar to *Reddit*, there are moderators but these are democratically elected by the community itself. Motivation for engaging with these platforms comes from a personal desire to assist colleagues, progress research, and receive recognition for it (Kubátová et al., 2012) – the same as that for peer review. Together, both have created open community-led collaboration and discussion platforms for their research disciplines.

3.3.2 Community-granted reputation and prestige

One of the key features of *Stack Exchange* is that it has an inbuilt community-based reputation system, karma, similar to that for *Reddit*. Identified peers rate or endorse the contributions of others and can indicate whether those contributions are positive (useful or informative) or negative. This provides a point-based reputation system for individuals based not just on the quantity of engagement with the platform and its peers alone, but also on the quality and relevance of those engagements, as assessed by the wider engaging community (`stackoverflow.com/help/whats-reputation`).
Peers have their status and moderation privileges within the platform upgraded as they gain reputation. Such automated privilege administration provides a strong social incentive for engaging within the community. Furthermore, peers who asked the original questions mark answers considered to be the most correct, thereby acknowledging the most significant contributions while providing a stamp of trustworthiness. This has the additional consequence of reducing the strain of evaluation and information overload for other peers by facilitating more rapid decision making, a behavior based on simple cognitive heuristics (e.g., social influences such as the "bandwagon effect" and position bias) (Burghardt et al., 2017). Threads can also be closed once questions have been answered sufficiently, based on a community decision, which enables maximum gain of potential karma points. This terminates further contribution but ensures that the knowledge is captured for future needs.

Therefore, karma and reputation are achieved and incentivized by building and contributing to a growing community and providing knowledgeable and comprehensible answers on a specific topic. Within this system, reputation points are distributed based on social activities that are akin to peer review, such as answering questions, giving advice, providing feedback, providing data, and generally improving the quality of work in the open. These points directly reflect an individual's contribution to that specific research community. Such processes ultimately have a very low barrier to entry, but also expose peer review to potential gamification through integration with a reputation engine, a social bias which proliferates through any technoculture (Belojevic et al., 2014).

3.3.3 Badge acquisition on Stack Overflow

An additional important feature of *Stack Overflow* is the acquisition of merit badges, which provide public stamps of group affiliation, experience, authority, identity and goal setting (Halavais et al., 2014). These badges define a way of locally and qualitatively differentiating between peers, and also symbolize motivational learning targets to achieve (Rughiniş and Matei, 2013). *Stack Overflow* also has a system of tag badges to attribute subject-level expertise, awarded once a peer achieves a certain voting score. Together, these features open up a novel reputation system beyond traditional measurements that is based on publications and citations of those articles and can also be used as an indication of expertise transferable beyond the platform itself. As such, a *Stack Exchange* model can increase the mobility of researchers who contribute in non-conventional ways (e.g., through software, code, teaching, data, art) and are based at non-academic institutes. There is substantial scope in creating a reputation platform that goes beyond traditional measurements to include social network influence and open peer-to-peer engagement. Ultimately, this can potentially transform the diversity of contributors to professional research and level the playing field for all types of formal contribution.

3.4 A GitHub-style model

Git is an open-source distributed version control system developed by the *Linux* community in 2005. *GitHub*, launched in 2008, works as a Web-based *Git* service and has become the de facto social coding platform for collaborative and open source development and code sharing (Kosner, 2012; Thung et al., 2013). It holds many potentially desirable features that might be transferable to a system of peer review (von Muhlen, 2011), such as its openness, version control and project management functionality, and system of accreditation and attribution for contributions. Despite its capability for not just sharing code, but also executable papers that automatically knit together text, data, and analyses into a living document, the true power of *GitHub* appears to be acknowledged infrequently by academic researchers (Priem, 2013).

3.4.1 Social functions of GitHub

Software review is an important part of software development, particularly for collaborative efforts. It is important that contributions are reviewed before they are merged into a code base, and *GitHub* provides this functionality. In addition, *GitHub* offers the ability to discuss specific issues, where multiple people can contribute to such a discussion, and discussions can refer to code segments or code changes and vice versa. *GitHub* also includes a variety of notification options for both users and project repositories. Users can watch repositories or files of interest and be notified of any new issues or commits (updates), and someone who has discussed an issue can also be notified of any new discussions of that same issue. Issues can also be tagged (labelled in a manner that allows grouping of multiple issues with the same tag), and assigned to one or more participants, who are then responsible for that issue. Another item that *GitHub* supports is a checklist, a set of items that have a binary state, which can be used to implement and store the status of a set of actions. *GitHub* also allows users to form organizations as a way of grouping contributors together to manage access to different repositories. All contributions are made public as a way for users to obtain merit.

Prestige at *GitHub* can be further measured quantitatively as a social product through the star-rating system, which is derived from the number of followers or watchers and the number of times a repository has been forked (i.e., copied) or commented on. This could ultimately shift the power dynamic in deciding what gets viewed and re-used away from editors, journals, or publishers to individual researchers. This then actually leverages a new mode of prestige, which is conferred through how work is engaged with by the wider community and not by the packaging in which it is contained (analogous to the prestige associated with journal brands).

Given these properties, it is clear that *GitHub* could be used to implement some style of peer evaluation and that it is well-suited to fine-grained iteration between reviewers and authors (Ghosh et al., 2012), given that all parties are identified. Making peer review a social process by distributing reviews to numerous peers, divides the burden and allows individuals to focus on their particular area of expertise. Peer review would operate more like a social network, with specific tasks (or repositories) being developed, distributed, and promoted through *GitHub*. As all code and data are supplied, peers would be able to assess methods and results comprehensively, which increases rigor, transparency, and replicability. Reviewers would also be able to claim credit and be acknowledged for their tracked contributions, and thereby quantify their impact on a project as a supply of individual prestige. This in turn facilitates an assessment of quality of reviews and reviewers. As such, evaluation becomes an interactive and dynamic process, with version control facilitating this all in a post-publication environment (Ghosh et al., 2012). The potential issue of proliferating non-significant work here is minimal, as projects that are not deemed to be in-teresting or of a sufficient standard of quality are simply never paid attention to in terms of follows, contributions, and re-use.

3.4.2 Current use of GitHub for peer review

An example use of *GitHub* for peer review already exists in *The Journal of Open Source Software* (*JOSS*, `joss.theoj.org`). *JOSS* provides a lightweight mechanism for software developers to quickly supplement their code with metadata and a descriptive paper, and then to submit this package for review and publication. *ReScience* (`rescience.github.io`) is another *GitHub*-based journal, created to publish replication efforts in computational science.

Here is a summary of how *JOSS* uses *GitHub*: The *JOSS* submission webpage converts a submission into a new *GitHub* issue of type "pre-review" in the *JOSS*-review repository (`github.com/openjournals/joss-reviews`). The editor-in-chief checks a submission and if deemed suitable for review, assigns it to a topic editor who in turn assigns it to one or more reviewers. The topic editor then issues a command that creates a new issue of type "review" with a check-list of required elements for the review. Each reviewer then performs their review by checking off elements of the review issue with which they are satisfied. When they feel the submitter needs to make changes to make an element of the submission acceptable, they can either add a new comment in the review issue, which the submitter will see immediately, or they can create a new issue in the repository where the submitted software and paper exist—which could also be on *GitHub*, but are not required to be—and reference said issue in the review. In either case, the submitter is automatically and immediately notified of the issue, prompting them to address the particular concern raised. This process can iterate repeatedly, as the goal of *JOSS* is not to reject submissions but to work with submitters until their submissions are deemed acceptable. If there is a dispute, the topic editor (as well as the main editor, other topic editors, and anyone else who chooses to follow the issue) can weigh in. At the end of this process, when all items in the review check-list are resolved, the submission is accepted by the editor and the review issue is closed. However, it is still available and is linked from the accepted (and now published) submission. A good future option for this style of model could be to develop host-neutral standards using Git for peer review. For example, this could be applied by simply using a prescribed directory structure, such as: `manuscript_version_1/peer_reviews`, with open commenting via the issues function.

3.5 A Wikipedia-style model

Wikipedia is the freely available, multi-lingual, expandable encyclopedia of human knowledge to all people. *Wikipedia*, like *Stack Exchange*, is another collective intelligence and authoring system whereby contributing communities are essentially unlimited in scope. Under a constant and instantaneous process of reworking and updat-

ing, new articles are added on a daily basis. *Wikipedia* operates through a system of collective intelligence based on linking knowledge workers through social media (Kubátová et al., 2012). Contributors to *Wikipedia* are largely anonymous volunteers, who are encouraged to participate mostly based on the principles guiding the platform (e.g., altruistic knowledge generation), and therefore often for reasons of personal satisfaction. Edits occur as cumulative and iterative improvements, and due to such a collaborative model, explicitly defining authorship becomes a complex task. Moderation and quality control is provided by a community of experienced editors and software-facilitated removal of mistakes, which can also help to resolve conflicts caused by concurrent editing by multiple authors (`wikipedia.org/wiki/Help:Edit_conflict`). Platforms already exist that enable multiple authors to collaborate on a single document in real time, including *Overleaf* and *Authorea*, which highlights the potential for this model to be extended into peer review. Communities of moderators *Wikipedia* also functionally exercise editorial power over content, and are nominated using conventional elections that variably account for their standing reputation. As such, the apparent "free for all" appearance of *Wikipedia* is actually more of a system of governance, based on implicitly shared values in the context of what is perceived to be useful for consumers, and transformed into operational rules to moderate the quality of content (Kelty et al., 2008).

3.5.1 "Peers" and "reviews" in a wiki-world

Wikipedia already has its own mode of peer review, requested by anyone as a way to receive ideas on how to improve articles that are already considered to be "decent" (`en.wikipedia.org/wiki/Wikipedia:Peer_review/guidelines`). It can be used for nominating potentially good articles that could become candidates for a featured article. Featured articles are considered to be the best articles *Wikipedia* has to offer, as determined by its editors and the fact that only ∼0.1% are featured, or an article of any grade. Users submitting a new request are encouraged to review an article from those already listed and encourage reviewers by replying promptly and appreciatively to comments. Compared to the conventional peer review process, where experts themselves participate in reviewing the work of another, the majority of the volunteers here, like most editors in *Wikipedia*, lack formal expertise in the subject at hand (Xiao and Askin, 2012). This is considered to be a positive thing within the *Wikipedia* community as it can make technically-worded articles more accessible to non-specialist readers.

This process clearly lacks the "peer" aspect of peer review, which can potentially lead to propagation of factual errors (e.g., Hasty et al. (2014)). This creates a general perception of low quality from the research community, in spite of difficulties in actually measuring this (Hu et al., 2007). However, as was originally the case with Open Access publishing, much of this perception can most likely be

explained by a lack of familiarity with the model, and we might expect comfort to increase and attitudes to change with increased engagement and understanding of the process (Xiao and Askin, 2014). If seeking expert input, users can invite editors from a subject-specific volunteers list or notify relevant WikiProjects. Furthermore, *Wikipedia* articles never "pass" a review, which, although part of the process of conventional validation, is of little actual value on the platform due to their dynamic nature. Indeed, wiki-communities appear to have distinct values to academic communities, being based more on inclusive community participation and mediation than on trust, exclusivity, and identification (Wang and Wei, 2012). Therefore, the process is perhaps best viewed as a process of "peer production", but where attainment of the level of peer is relatively lower to that of an accredited expert. This provides a difference in community standing for *Wikipedia* content, with value being conveyed through contemporariness, mediation of debate, and transparency of information, rather than any perception of authority as with traditional scholarly works (Black, 2008). Such a process could be feasibly combined with trust metrics for verification, developed for sociology and psychology to describe the relative standing of groups or individuals in virtual communities (`en.wikipedia.org/wiki/Trust_metric`).

3.5.2 Democratization of peer review

The advantage of *Wikipedia* over traditional review-then-publish processes comes from the fact that articles are enhanced consistently as new articles are integrated, statements are reworded, and factual errors are corrected as a form of iterative bootstrapping. Therefore, while one might consider a *Wikipedia* page to be of insufficient quality relative to a peer reviewed article at a given moment in time, this does not preclude it from meeting that quality threshold in the future. Therefore, *Wikipedia* might be viewed as an information trade-off between accuracy and scale, but with a gap that is consistently being closed as the overall quality improves. Another major statement that a *Wikipedia*-style of peer review makes is that rather than being exclusive, it is an inclusive process that anyone is allowed to participate in and the barriers to entry are very low—anyone can potentially be granted peer status and participate in the debate and vetting of knowledge. Therefore, *Wikipedia* represents a fairly extreme alternative to peer review where traditionally the barriers to entry are very high (based on expertise), to one where the pool of potential peers is relatively large (Kelty et al., 2008). This represents an enormous shift from the generally technocratic process of conventional peer review to one that is inherently more democratic. However, while the number of contributors is very large, more than 30 million, one third of all edits are made by only 10,000 people, just 0.03% (`en.wikipedia.org/wiki/Wikipedia:List_of_Wikipedians_by_number_of_edits`). This is broadly similar to what is observed in current academic peer review systems, where the majority of the work is performed by a minority of the participants

(Kovanis et al., 2016; Gropp et al., 2017; Fox et al., 2017). Any wiki-based peer review system would also alleviate the increasing burden on editors by distributing the endeavor more efficiently among members of the wider community—a high-risk, high-gain approach to generating academic capital (Black, 2008). A possible risk is the creation of a highly conservative network of norms due to the governance structure, which could end up being even more bureaucratic and creating community silos rather than coherence (Heaberlin and DeDeo, 2016). To date, attempts at implementing a *Wikipedia*-like editing strategy for journals have been largely unsuccessful to date (e.g., at *Nature* (Zamiska, 2006)). There are intrinsic differences in authority models used in *Wikipedia* communities (where the validity of the end result derives from verifiability, not personal authority of authors and reviewers) that would need to be aligned with the norms and expectations of science communities. In the latter, author statements and peer reviews are considered valid because of the personal, identifiable status and reputation of authors, reviews an editors, which could be feasibly combined with *Wikipedia* review models into a single solution. However, a more rigorous editorial review process is the reason why the original form of *Wikipedia*, known as *Nupedia*, ultimately failed (Sanger, 2005). Future developments of any *Wikipedia*-like peer review tool could expect strong resistance from universities due to potential disruption to assessment criteria, funding assignment, and intellectual property, as well as from commercial publishers since academics would be releasing their research to the public for free instead of them.

3.6 A Hypothesis-style annotation model

Hypothesis is a lightweight, portable Web annotation tool that operates across publishing platforms (Perkel, 2015), ambitiously described as a "peer review layer for the entire Internet" (Farley, 2011). It relies on pre-existing published content to function, similar to other annotation services, such as *PubPeer* and *PaperHive*. Annotation is a process of enriching research objects through the addition of knowledge, and also provides an interactive educational opportunity through raising questions and creating opportunities to collect the perspectives of multiple peers in a single venue; thus possessing a dual functionality for collaborative reading and writing. Web annotation services like *Hypothesis* allow the annotations (such as comments or peer reviews) to live alongside the content but also separate from it, allowing communities to form and spread across the internet and across content types, such as HTML, PDF, EPUB, or other formats (Whaley, 2017). Further, as of February 2017, annotation became a Web standard recognized by the Web Annotation Working Group, W3C (2017). Under this model of Web annotation described by the W3C, annotations belong to and are controlled by the user rather than any individual publisher or content host. Users use a bookmarklet or browser extension to annotate any webpage they wish, and form a community of Web citizens.

Hypothesis permits the creation of public, group private, and individual private annotations, and is therefore compatible with a range of open and closed peer review models. Web annotation services not only extend peer review from academic and scholarly content to the whole Web but open up the ability to annotate to any Web-browser. While the platform concentrates on focus groups within publishing, journalism, and academia, *Hypothesis* offers a new way to enrich, fact check, and collaborate on online content. Unlike *Wikipedia*, the core content never changes but the annotations are viewed as an overlay service on top of static content. This also means that it can occur at any time during the publishing process, including the pre-print stage. Reviewers often provide annotated versions of submitted manuscripts during conventional peer review, and Web annotation is part of the digitization of this process, while also decoupling it from journal hosts. A further benefit of Web annotations is that they are precise, since they can be applied in line rather than at the end of an article as is the case with formal commenting.

Annotations have the potential to enable new kinds of workflows where editors, authors, and reviewers all participate in conversations focussed on research manuscripts or other digital objects, either in a closed or public environment (Vitolo et al., 2015). At the present, activity performed by *Hypothesis* and other Web annotation services is poorly recognized in scholarly communities, although such activities can be tied to *ORCID*. However, there is definite value in such services like *PubPeer*, an online community mostly used for identifying cases of academic misconduct and fraud, which is perhaps best known for its users-led post-publication critique of a *Nature* paper on STAP (Stimulus-Triggered Acquisition of Pluripotency) cells. This ultimately prompted the formal retraction of the paper demonstrating that post-publication annotation and peer review, as a form of self-correction and fraud detection, can out-perform that of the conventional pre-publication process. *PubPeer* has also been leveraged as a way for mass-reporting post-publication checks for the soundness of statistical analyses. One large-scale analysis using a tool called *statcheck* was used to post 50,000 annotations on the psychological literature (Singh Chawla, 2016) as a form of large-scale public audit for published research.

3.7 A blockchain-based model

Peer review has the potential to be reinvented as a more efficient, fair, and otherwise attribute-enabled process through blockchains, a computer data structure that operates a distributed public ledger. A blockchain connects a row of data blocks through a cryptographic function, with each block containing a time stamp and a link to the previous block in the chain. This system is decentralized, distributed, immutable, and transparent (Nakamoto, 2008; Antonopoulos, 2014; Yli-Huumo et al., 2016). Perhaps most importantly, individual chains are managed

by peer-to-peer networks that collectively adhere to specific validation protocols. Blockchain became widely known as the data structure in Bitcoin due to its ability to efficiently record transactions between parties in a verifiable and permanent manner. It has also been applied to other uses including sharing verified business transactions, proof of ownership of legal documents, and distributed cloud storage.

The blockchain technology could be leveraged to create a tokenized peer review system involving penalties for members who do no uphold the adopted standards and vice versa. A blockchain-powered peer-reviewed journal could be issued as a token system to reward contributors, reviewers, editors, commentators, forum participants, advisors, staff, consultants, and indirect service providers involved in scientific publishing (Swan, 2015). Such rewards could be in the form of reputation and/or remuneration, potentially through a form of digital currency (say *Science Coins*). Through a system of community trust, blockchains could be used to handle the following tasks:

1. Authenticate scientific papers (using time stamps and checksums), combating fraudulent science;

2. Permit and encourage reviewers to actively engage in the scientific community;

3. Reward reviewers for peer reviews with Science Coins;

4. Allow authors to contribute by giving Science Coins;

5. Support verification and replicability of research.

6. Keeping reviewers and authors anonymous, while providing a validated certification of their identity as researchers, and rewarding them.

This could help to improve the quality and responsiveness of peer reviews, as these are published publicly and the different participants are rewarded for their contributions. For instance, reviewers for a blockchain-powered peer-reviewed journal could invest tokens in their comments and get rewarded if the comment is upvoted by other reviewers and the authors. All tokens need to be spent in making comments or upvoting other comments. When the peer review is completed, reviewers get rewarded according to the quality of their remarks. In addition, the rewards can be attributed even if reviewers and authors identity is kept secret; such a system can decoupled the quality assessment of the reviews from the reviews themselves, such that reviewers get credited while their reviews are kept anonymous. Moreover, increased transparency and interactions are facilitated between authors, reviewers, the scientific community, and the public. The journal *Ledger*, launched in 2015, is the first academic journal that makes use of a system of digital signatures and time stamps based on blockchain technology (`ledgerjournal.org`). The aim is to generate irrevocable proof that a given manuscript existed on the date of publication.

Furthermore, blockchain-based models offer the potential to go well beyond peer review, possibly integrating all functions of publication in general. They could be used to support data publication, research evaluation, incentivization, and research fund distribution. A relevant example is a proposed decentralized peer review group as a way of managing quality control in peer review via blockchain through a system of cohort-based training (Dhillon, 2016). This has also been leveraged as a "proof of existence" platform for scientific research (Torpey, 2015) and medical trials (Carlisle, 2014). However, the uptake from the academic community remains low thus far, despite claims that it could be a potential technical fix to the reproducibility crisis in research (Bartling and Fecher, 2016). As with other novel processes, this is likely due to broad-scale unfamiliarity with blockchain, and perhaps even discomfort due to its financial association with Bitcoin.

3.8   AI-assisted peer review

Another frontier is the advent of machine learning (ML) and neural network tools that may potentially assist with the peer review process. Machine learning, as a technique, is rapidly becoming a service that can be utilized at a low cost by an increasing number of individuals. For example, *Amazon* now provides ML as a service through their *Amazon Web Services* platform, *Google* released their ML framework, *TensorFlow*, and *Facebook* have similarly contributed code of their *Torch* scientific learning framework. ML has been very widely adopted in tackling various challenges, including image recognition, content recommendation, fraud detection, and energy optimization. In higher education, adoption has been limited to automated evaluation of teaching and assessment, and in particular for plagiarism detection. The primary benefits of Web-based peer assessment are limiting peer pressure, reducing management workload, increasing student collaboration and engagement, as well as improving the understanding of peers as to what critical assessment procedures involve (Li et al., 2009).

The same is approximately true for using computer-based automation for peer review, for which there are three main practical applications. The first is determining whether a piece of work under consideration meets the minimal requirements of the process to which it has been submitted (i.e., for recommendation). For example, does a clinical trial contain the appropriate registration information, are the appropriate consent statements in place, have new taxonomic names been registered, and does the research fit in with the existing body of published literature (Sobkowicz, 2008)? The computer might also look at consistency through the paper (for example searching for statistical error or method description incompleteness: if there is a multiple group comparison, is the p-value correction algorithm indicated?) This might be performed using

a simpler text mining approach, as is performed by *statcheck* (Singh Chawla, 2016). Under normal technical review these criteria need to be (or should be) checked manually either at the editorial submission stage or at the review stage. ML techniques can automatically scan documents to determine if the required elements are in place, and can generate an automated report to assist review and editorial panels, facilitating the work of the human reviewers. Moreover, any relevant papers can be automatically added to the editorial request to review, enabling referees to automatically have a greater awareness of the wider context of the research. This could also aid in preprint publication before manual peer review occurs.

The second approach is to automatically determine the most appropriate reviewers for a submitted manuscript by using a co-authorship network data structure (Rodriguez and Bollen, 2008). The advantage of this is that it opens up the potential pool of referees beyond whom is simply known by an editor or editorial board. Removing human-intervention from this part of the process reduces potential biases (e.g., author recommended exclusion or preference) and can automatically identify potential conflicts of interest (Khan, 2012). Dall'Aglio (2006) suggests ways this algorithm could be improved, for example through cognitive filtering to automatically analyze text and compare that to editor profiles as the basis for assignment. This could be built upon for referee selection by using an algorithm based on social networks, which can also be weighted according to the influence and quality of participant evaluations (Rodriguez et al., 2006), and referees can be further weighted based on their previous experience and contributions to peer review and their relevant expertise, thereby providing a way to train and develop the identification algorithm.

Finally, given that machine-driven research has been used to generate substantial and significant novel results based on ML and neural networks, we should not be surprised if, in the future, they can have some form of predictive utility in the identification of novel results during peer review. In such a case, machine learning would be used to predict the future impact of a given work (e.g., future citation counts), and in effect to do the job of impact analysis and decision making instead of or alongside a human reviewer. We have to keep a close watch on this potential shift in practice as it comes with obvious potential pitfall by encouraging even more editorial selectivity, especially when network analysis is involved. For example, research in which a low citation future is predicted would be more susceptible to rejection, irrespective of the inherent value of that research. Conversely, submissions with a high predicted citation impact would be given preferential treatment by editors and reviewers. Caution in any pre-publication judgements of research should therefore always be adopted, and not be used as a surrogate for assessing the real world impact of research through time. Machine learning is not about providing a total replacement for human input to peer review, but more how different tasks could be delegated or refined through automation.

Some platforms already incorporate such methods for a variety of purposes. *Scholastica* (`scholasticahq.com`) includes real-time journal performance analytics that can be used to assess and improve the peer review process. *Elsevier* uses a system called Evise (`elsevier.com/editors/evise`) to check for plagiarism, recommend reviewers, and verify author profile information by linking to *Scopus*. The *Journal of High Energy Physics* uses automatic assignment to editors based on a keyword-driven algorithm (Dall'Aglio, 2006). This process has the potential to be entirely independent from journals and can be easily implemented as an overlay function for repositories, including pre-print servers. As such, it can be leveraged for a decoupled peer review process by combining certification with distribution and communication. It is entirely feasible for this to be implemented on a system-wide scale with researcher databases such as *ORCID* becoming increasingly widely adopted. However, as the scale of such an initiative increases, the risk of over-fitting also increases due to the inherent complexity in modelling the diversity of research communities, although there are established techniques to avoid this. Questions have been raised about the impact of such systems on the practice of scholarly writing, such as how authors may change their approach when they know their manuscript is being evaluated by a machine (Hukkinen, 2017), or how machine assessment could discover unfounded authority in statements by authors through analysis of citation networks (Greenberg, 2009). One additional potential drawback of automation of this sort is the possibility for detection of false positives that might discourage authors from submitting.

Finally, it is important to note that ML and neural networks are largely considered to be conformist, so they have to be used with care (Szegedy et al., 2014), and perhaps only for recommendations rather than decision making. The question is not about whether automation produces error, but whether it produces less error than a system solely governed by human interaction. And if it does, how does this factor in relation to the benefits of efficiency and potential overhead cost reduction? Nevertheless, automation can potentially resolve many of the technical issues associated with peer review and there is great scope for increasing the breadth of automation in the future. Initiatives such as *Meta*, an AI tool that searches scientific papers to predict the trajectory of research (`meta.com`), highlight the great promise of artificial intelligence in research and for application to peer review.

3.9 Peer review for non-text products

Peer review has also evolved when used outside of traditional text-based scholarly publications to a wider variety of research outputs, policies, processes, and even people. These non-text products are increasingly being recog-

nized as intellectual contributions to the research ecosystem. In order for the creators (authors) of these products to receive academic credit, they must currently be integrated into the publication system that forms the basis for academic assessment and evaluation. Peer review of methodologies, such as *protocols.io* (`protocols.io`), allows for detailed OPR of methods while also promoting reproducibility and refinement of techniques. This can help other scholars to begin work on related projects and test methodologies due to the openness of both the protocols themselves and the comments on them (Teytelman et al., 2016). Digital humanities projects, which include visualizations, text processing, mapping, and many other varied outputs, have been a subject for re-evaluating the role of peer review, especially for the purpose of tenure and evaluation (Ball et al., 2016). In 2006, the *Modern Languages Association* released a statement on the peer review and evaluation of new forms of scholarship, insisting that they "be assessed with the same rigor used to judge scholarly quality in print media" (Stanton et al., 2007). Fitzpatrick (2011a) also considered the idea of an objective evaluation of non-text products in the humanities as well as the challenges of evaluation of a digital product that may have much more to review than a traditional text product, including community engagement and sustainability practices. To work with these non-text products, humanities scholars have used multiple methods of peer review and embraced OPR in order to adapt to the increased creation of non-text, multimedia scholarly products and to integrate these products into the scholarly record and review process (Anderson and McPherson, 2011).

3.9.1 Software peer review

Software represents another area where traditional peer review has evolved. In software, peer review of code has been a standard part in computationally-intensive research for many years, particularly as a post-software creation check. Additionally, peer-programming (also known as pair-programming) has been growing in popularity, especially as part of the Agile methodology, where it is employed as a check made during software creation (Lui and Chan, 2006). Software development and sharing platforms, such as *GitHub*, support and encourage social code review, which can be viewed as a form of peer review that takes place both during creation and afterwards. However, developed software has not traditionally been considered an academic product for the purpose of hiring, tenure, and promotion. Likewise, this form of evaluation has not been formally recognized as peer review by the academic community yet.

When it comes to software development, there is a dichotomy of review practices. On the one hand, software developed in open source communities (not all software is released as open source; some is kept as proprietary for commercial reasons) relies on peer review as an intrinsic part of its existence, from creation and through continual evolution. On the other hand, software created in academia is typically not subjected to the same level of scrutiny. At present, there is no requirement for software, used to produce academic publications, to be released as part of the publication process, let alone be closely checked as part of the review process, though this may be changing due to government mandates and community concerns about reproducibility. One example from Computer Science is ACM SIGPLAN's Programming Language Design and Implementation conference that encourages the submission of supporting material (including code) for review by a separate technical committee. Papers with successfully evaluated artifacts get stamped with seals of approval visible in the conference proceedings. ACM is adopting a similar strategy on a wider scale through its Task Force on Data, Software, and Reproducibility in Publication (`acm.org/data-software-reproducibility`). Academic code is sometimes released as open source and many such released codebases have led to remarkable positive changes, with prominent examples including the Berkeley Software Distribution (BSD), upon which the Mac operating system (MacOS) is built; the ubiquitous TCP/IP Internet protocol stack; the Squid web proxy; the Xen hypervisor, which underpins many cloud computing infrastructures; Spark, the big data stream processing framework; and the Weka machine learning suite.

In order to gain recognition for their software work, authors initially made as few changes to the existing system as possible and simply wrote traditional papers about their software, which became acceptable in an increasing number of journals over time (see the extensive list compiled by the UK's Software Sustainability Institute: `software.ac.uk/which-journals-should-i-publish-my-software`). At first, peer review for these software articles was the same as for any other paper, but this is changing now, particularly as journals specializing in software (e.g., *SoftwareX* (`journals.elsevier.com/softwarex`), the *Journal of Open Research Software* (*JORS*, `openresearchsoftware.metajnl.com`), the *Journal of Open Source Software* (*JOSS*, `joss.theoj.org`)) are emerging. The material that is reviewed for these journals is both the text and the software. For *SoftwareX* (`elsevier.com/authors/author-services/research-elements/software-articles/original-software-publications#submission`) and *JORS* (`openresearchsoftware.metajnl.com/about/#q4`), the text and the software are reviewed equally. For *JOSS*, the review process is more focused on the software (based on the *rOpenSci* model (Ross et al., 2016) and less on the text, which is intended to be minimal (`joss.theoj.org/about#reviewer_guidelines`). The purpose of the review also varies across these journals. In *SoftwareX* and *JORS*, the goal of the review is to decide if the paper is acceptable and to improve it through a non-public editor-mediated iteration with the authors and the anonymous reviewers. While in *JOSS*, the goal is to accept most papers after improving them if needed, with the reviewers and authors ideally

communicating directly and publicly through *GitHub* issues. Although submitting source code is still not required for most peer review processes, attitudes are slowly changing. As such, authors increasingly publish works presented at major conferences (which are the main channel of dissemination in computer science) as open source.

3.9.2 Data peer review

Many journals in biological science already ask authors to release data to reviewers when they submit a paper (e.g., *Journal of Cell Science*, *Microbial Genomics*, *Royal Society Open Science*), and journals can ask reviewers to say whether or not they have reviewed the data themselves. Making data available to reviewers and editors can help to correct obvious errors and, more importantly, serves as a strong incentive for authors to check data and analyses thoroughly before releasing them. For example, errors in datasets, improper data archiving standards, or simply a reluctance to release datasets can play a role in whether a particular journal rejects a paper. If journals enforced this more strictly, an improvement in data archiving standards would likely be the consequence. However, some reviewers, who may already be overburdened, might feel that checking data accuracy while reviewing the manuscript at the same time and receiving only little to no reward for their work (see §2.2) is unfair. The Peer Reviewers' Openness Initiative (`opennessinitiative.org/the-initiative`) states that all data should be made publicly available for the purposes of evaluation and reproduction and indicates that there is wider scope for the development of data peer review in the future (Morey et al., 2016).

3.10 Using multiple peer review models

While individual publishers may use specific peer review methods when peer review is controlled by the author of the document to be reviewed, multiple peer review models can be used either in series or in parallel. For example, the *FORCE11* Software Citation Working Group used a series of three different peer review models and methods to iteratively improve their principles document, leading to a journal publication (Smith et al., 2016). Initially, the document that was produced was made public and reviewed by *GitHub* issues (`github.com/force11/force11-scwg` [§3.4]). The next version of the document was placed on a website, and new reviewers commented on it both through additional *GitHub* issues and through *Hypothesis* (`via.hypothes.is/https://www.force11.org/software-citation-principles` [§3.6]). Finally, the document was submitted to *PeerJ Computer Science*, which used a pre-publication review process that allowed reviewers to sign their reviews and the reviews to be made public along with the paper authors' responses after the final paper was accepted and published (Klyne, 2016; Kuhn, 2016a,b). The authors also included an appendix that summarized the reviews and responses from the second phase. In summary, this document underwent three sequential and non-conflicting review processes and methods, where the second one was actually a parallel combination of two mechanisms.

Using such hybrid evaluation methods can prove to be quite successful not just for reforming the peer review process but also to improve the impact of scientific publications in different ways. One could envision such a hybrid system with elements from the different models we discussed. *Stack Overflow* helps to surface practical solutions and their trade-offs. As such, it is well geared for research in practice, collating the community's knowledge through collective validation, comparison, and innovation. *Reddit*-style discussion threads, which are characterized as being general and loose (not focused), are very effective for public engagement and for hosting virtual conferences or parallel discussion forums augmenting "physical" conferences. It might be necessary to adjust the *Reddit* karma system to give more weight to comments rather than new posts to reduce the incentive to publish frequent, yet potentially low quality work.

4 A hybrid peer review platform

In Section 3, we summarized the positive and negative traits of a range of individual existing social platforms. Each of these traits can be applied to address specific social or technical criticisms of conventional peer review, as outlined in Section 2. Many of them are overlapping and can be modeled into, and leveraged for, a single hybrid platform. The advantage here is that they each relate to the core non-independent features required for any modern peer review process or platform: *quality control*, *certification*, and *incentivization*. Only by harmonizing all three of these, while grounding development in diverse community stakeholder engagement, can the implementation of any future model of peer review be ultimately successful. Such a system has the potential to greatly disrupt the current coupling between peer review and journals, and lead to an overhaul of scholarly communication to become one that is fit for the modern scholarly research environment.

4.1 Quality control and moderation

Quality control is the core function of peer review. Typically, this has been administered in a closed system, where editorial management formed the basis. A strong coupling of peer review to journals plays an important part in this due to the association of researcher prestige with journal brand. By looking at platforms such as *Wikipedia* and *Reddit*, it is clear that community self-organization and governance represent a possible alternative when combined with a core community of moderators. These moderators would have the same operational functionality as editors in terms of gatekeeping and facilitating the process of engagement, but combined with the role of a Web forum moderator. Research communities could elect groups of moderators based on expertise, prior engagement with peer review,

and transparent assessment of their reputation. This layer of moderation could be fully transparent in terms of identity by using persistent identifiers such as *ORCID*. Different communities could have different norms and procedures here to govern content and engagement, and self-organize into individual but connected platforms, similar to *Stack Exchange* or *Reddit*. *ORCID* has a further potential role of providing the possibility for a public archive of researcher information and metadata (e.g., publishing histories) that can be leveraged using automated techniques to match potential referees to items of interest, while avoiding conflicts of interest.

In such a system, published objects could be pre-prints, data, code, or any other digital research output. If these are combined with management through version control, similar to *GitHub*, quality control is provided by having a system of automated but managed invited review, public interaction and collaboration (like with *Stack Exchange*), and transparent refinement. This would also help to prevent a "the rich get richer" scenario, as semi-automation ensures that all content has the same chance of being interacted with. Engagement could be conducted via a system of issues and public comments, as on *GitHub*, where the process is not to reject submissions, but provide a system of constant improvement. Such a system is already implemented successfully at *JOSS*. Both community moderation and crowd sourcing would play an important role here to prevent underdeveloped feedback that is not constructive and could delay efficient manuscript progress. This could be further integrated with a blockchain process so that each addition to the process is transparent and verifiable.

When authors and moderators deem the review process to have been sufficient for an object to have reached a community-decided level of quality or acceptance, threads can be closed (but remain public with the possibility of being re-opened, similar to *GitHub* issues), indexed, and the latest version is assigned a persistent identifier, such as a *CrossRef* DOI, as well as an appropriate license. If desired, these objects could then form the basis for submissions to journals, perhaps even fast-tracking them as the communication and quality control would already have been completed. Such a process would promote inclusive participation, community interaction, and quality would become a function of how information is engaged with, digested, and reused. The role of peer review, then, would be coupled with the concept of a "living published unit", but one that is independent of journals themselves. Indeed, the role of journals and publishers would hence be dependent on how well they justify their added value once community-wide and public dissemination and peer review have been decoupled from them.

4.2 Certification and reputation

The current peer review process is generally poorly recognized as a scholarly activity. It remains quite imbalanced between publishers who receive financial gain for organising it and researchers who receive little or no compensation for performing it. Opacity in peer review provides a way for others to capitalize on the process, as this provides a mechanism for those managing it, rather than performing it, to take credit in one form or another. This, at least in part, explains the resistance from many publishers in providing any form of substantive recognition to peer reviewers. Exposing the process, decoupling it from journals (see previous subsection), and providing appropriate recognition to those involved, helps to return peer review to its synergistic, intra-community origin. Performance metrics provide a way of certifying the peer review process as well as the basis for incentivizing engagement. As outlined above, a fully transparent and interactive process of engagement combined with reviewer identification exposes the level of engagement and the added value from each participant.

Certification can be provided to referees based on their engagement with the process: community evaluation of their contributions (.e.,g *Amazon*, *Reddit*, or *Stack Exchange*), combined with their reputation as authors. Rather than having anonymous or pseudonymous participants, for peer review to work well it would require full identification to connect on-platform reputation and authorship history. Therefore, rather than a journal-based form, certification is granted based on continuing engagement with the research process and is revealed at the article (or object) and individual level. Communities would need to decide whether or not to set engagement filters based on quantitative measures of experience or reputation, and what this should be for different activities (e.g., as employed at *ScienceOpen*). This should be highly appealing not just to researchers, but also to those in charge of hiring, tenure, promotion, grant funding, and research assessment, and therefore could become an important factor in future policy development. Models like *Stack Exchange* are ideal candidates for such a system because achievement of certification takes place via a process of community engagement and can be quantified through a simple and transparent up-voting and down-voting scheme, combined with achievement badges. Any outputs from assessment could be portable and applied to *ORCID* profiles, external webpages, and continuously updated and refined through further activity. While a star system does not seem appealing due to the inherent biases associated with it, this quantitative way of "reviewing the reviewers" creates a form of dynamic social reputation. As this is decoupled from journals, it alleviates all of the well-known issues with journal-based ranking systems and is fully transparent. By combining it with moderation, as outlined above, gaming is also prevented (e.g., by providing numerous low quality engagements). Integrating a blockchain-based token system could also reduce potential for such gaming. Most importantly here, it is research communities, and engagement within them, that form the basis of certification, and reputation should evolve continuously with this.

4.3 Incentives for engagement

Incentives are required to motivate and encourage wider participation and engagement with peer review. As such, this requires lowering the threshold of entry for different research communities. The most widely-held reason for performing peer review is a sense of academic altruism or duty to the research community. However, at present this is imbalanced and researchers still receive far too little credit as a way of recognizing their efforts. This is directly tied to certification and reputation, as above, which is the ultimate goal of any incentive system.

New ways of incentivizing peer review can be developed by quantifying engagement with the process and tying this in to academic profiles, such as *ORCID*. To some extent this is already performed via *Publons*, where the records of individuals reviewing for a particular journal can be integrated into *ORCID*. This could easily be extended to include aspects from *Reddit*, *Amazon*, and *Stack Exchange*, where participants receive virtual rewards, such as points or karma, for engaging with peer review and having those activities further evaluated and ranked by the community. After a certain quantified threshold has been achieved, a hierarchical award system could be developed into this, and then be subsequently integrated into *ORCID*. Such awards or badges could include "Top reviewer", "Verified reviewer", "Community leader'," or whatever individual communities decide is best for them. This can form an incentive loop, where additional engagement abilities are acquired based on achievement of such badges.

Highly-rated reviews gain more exposure and more credit, thus there incentive is to engage with the process in a way that is most beneficial to the community. Engagement with peer review and community evaluation of that then becomes part of a verified academic record, which can then be used as a way of establishing individual prestige. Such a system would be automatically integrated with any published content itself and objects could be similarly granted badges, such as "Community reviewed," "Community accepted," or "500 upvotes" as a way of quantifying the process. Therefore, there would be a dual incentive for authors to maximize engagement from the research community and for that community to productively engage with content. A potential extension of this in the form of monetization (e.g., through a blockchain protocol) is perhaps unwise at it may lead to a potential distortion of incentives.

4.4 Challenges

None of the ideas proposed here are particularly radical, representing more the recombination of existing variants that have succeeded or failed to varying degrees. A key challenge that our proposed hybrid system will have to overcome is simultaneous uptake across the whole scholarly ecosystem. In particular this proposed system involves a requirement for standardised communication between a range of actors. Real shifts will occur where elements of this system can be taken up by specific communities, but remain interoperable between them. Identifying sites where stepwise changes in practice are desirable to a community is an important next step. However, it is clear that recent advances in technology can play a significant role in systemic changes to peer review. High quality implementations of these ideas in systems that communities can choose to adopt may act as *de facto* standards that help to build towards consistent practice and adoption.

One aspect that we did not examine here in detail is the use of instant messaging services, like *Slack* or *Gitter*. These are widely used for project communication and operate analogous to a real-time collaboration system with instantaneous and continuous "peer review". Therefore, while such activities can be used to supplement other hybrid platforms, the concept as an independent or stand-alone mode of peer review is quite distant from the other models we discuss here.

5 Conclusions

If the current system of peer review were to undergo peer review, it would undoubtedly achieve a "revise and resubmit" decision. As Smith (2010) succinctly stated, "we have little or no evidence that peer review 'works,' but we have lots of evidence of its downside". The Internet has changed our expectations of how communication works, and enabled a wide array of new, technologically-enabled possibilities to change how we communicate and interact online. Peer review has also recently become an online endeavor, but few organizations who conduct peer review have adopted Internet-style communication norms. This leaves a gap in what is possible with current technology and social norms and what we are doing to ensure the reliability and trustworthiness of published science. Peer review is a critical part of an effective scientific enterprise, but many of those who conduct peer review and depend upon it do not fully understand the theoretical and empirical basis for it. This means that our efforts to advance and change peer review are being driven by organizational goals such as market position and profit, and not by the needs of academia.

Existing, popular online communication systems and platforms were designed to attract a huge following, not to ensure the ethics and reliability of effective peer review. Numerous front-end Web applications already implement all of the essential core traits for creating a widely distributed, diverse peer review ecosystem. We already have the technology we need. However, it will take a lot of work to integrate new technology-mediated communication norms into effective, widely-accepted peer review models, and connect these together seamlessly so that they become inter-operable as part of a sustainable scholarly communications infrastructure. Identity is a core factor driving online communication adoption and social norms and practices of current peer review – both

how it is traditionally conducted with editorial management, and what will be possible with novel models online.

These socio-technological barriers cannot be overcome by simply creating platforms and expecting researchers to use them. Rather, as others have suggested (e.g., Moore et al. (2017); Prechelt et al. (2017)), platforms should be developed with community engagement, education, and capacity building as core traits, in order to help understand the cultural processes and needs of different disciplines and create solutions around those. Coordinated efforts are required to teach and market the purpose of peer review to researchers. More effective engagement is clearly required to emphasize the distinction between the idealized processes of peer review along with the perceptions and applications of it and the resulting products and services available to conduct it. These would help to close the divergence between the social ideology and the technological application of peer review.

In this paper, we present an overview of what the key features of a hybrid, integrated peer review and publishing platform might be and how these could be combined. These features are embedded in research communities, which can not only set the rules of engagement but also form the judge, jury, and executioner for quality control, moderation, and certification. The major benefit of such a system is that peer review becomes an inherently social and community-led activity, decoupled from a traditional journal-based system, and instead part of the commons. The "Principle of Maximum Bootstrapping" outlined by Kelty et al. (2008) is highly congruent with this social ideal for peer review, where new systems are based on existing communities of expertise, quality norms, and mechanisms for review. Diversifying peer review in such a manner is an intrinsic part of a system of reproducible research (Munafò et al., 2017). Making use of persistent identifiers such as *DataCite*, *CrossRef*, and *ORCID* will be essential in binding the social and technical aspects of this to an interoperable, sustainable and open scholarly infrastructure (Dappert et al., 2017).

However, we recognize that any technological advance is rarely innocent or unbiased, and while Web 2.0 technologies open up the possibility for increased participation in peer review, it would still not be inherently democratic (Elkhatib et al., 2015). As Belojevic et al. (2014) remark, when considering tying reputation engines to peer review, we must be aware that this comes with implications for values, norms, privilege and bias, and the industrialization of the process (Lee et al., 2013). Peer review is socially and culturally embedded in scholarly communities and has an inherent diversity in values and processes, which we must have a deep awareness of and appreciation for. Evidence-based research on peer review itself would help to build our collective understanding of the process and guide the design of ad-hoc solutions (Drummond, 2016). Further research should also focus on the challenges faced by researchers from peripheral nations, particularly for those who are non-native English

speakers, and increase their influence as part of the globalization of research (Salager-Meyer, 2008, 2014; Fukuzawa, 2017). The scholarly publishing industry could help to foster such research by starting to share its data on peer review (Squazzoni et al., 2017), with their incentive being to help improve the process.

Academics have been entrusted with an ethical imperative towards accurately generating, transforming, and disseminating new knowledge through peer review and scholarly communication. Peer review started out as a collegial discussion between authors and editors. Since this humble origin, it has vastly increased in complexity and become systematized and commercialized in line with the neo-liberal evolution of the modern research institute. This system is proving to be a vast drain upon human and technical resources due to the increasingly unmanageable workload involved in scholarly publishing. There are lessons to learn from the Open Access movement, which started as a set of principles by people with good intentions, but was subsequently converted into a messy system of mandates, policies, and increased costs that is becoming increasingly difficult to navigate. Commercialization has inhibited the progress of scholarly communication, and can no longer keep pace with the generation of new ideas in a digital world.

The research community has the opportunity to help create an efficient and socially-responsible system of peer review. The history, technologies, and social justification to do so all exist. Research communities need to embrace the opportunities gifted to them and work together across stakeholder boundaries (e.g., with research funders, libraries, professional communicators) to create a more optimal system of peer review aligned with the diverse needs of non-independent research communities. By decoupling peer review, and with it scholarly communication, from commercial entities and journals, it is possible to return it to the core principles upon which it was founded more than a century ago. Through this, knowledge generation and access can become a democratic process and academics can fulfill the criteria that has been entrusted to them as the creators and guardians of knowledge.

Author contributions
All authors contributed to this document via the Overleaf collaborative writing platform. Author order was set in order of first in, first listed.

Competing interests
JPT works for ScienceOpen; TRH works for OpenAIRE.

Grant information

References

Adam, D. (2010). Climate scientists hit out at 'sloppy' melting glaciers error. *The Guardian*, https://www.theguardian.com/environment/2010/feb/08/climate-scientists-melting-glaciers. Accessed: 2017-06-12.

Al-Rahawi, I. b. A. (c900). *Practical Ethics of the Physician (Adab al-Tabib)*.

Albert, A. Y., Gow, J. L., Cobra, A., and Vines, T. H. (2016). Is it becoming harder to secure reviewers for peer review? a test with data from five ecology journals. *Research Integrity and Peer Review*, 1(1):14, DOI: 10.1186/s41073-016-0022-7.

Alvesson, M. and Sandberg, J. (2014). Habitat and habitus: Boxed-in versus box-breaking research. *Organization Studies*, 35(7):967–987, DOI: 10.1177/0170840614530916.

Anderson, S. and McPherson, T. (2011). Engaging digital scholarship: Thoughts on evaluating multimedia scholarship. *Profession*, pages 136–151, DOI: 10.1632/prof.2011.2011.1.136.

Antonopoulos, A. M. (2014). *Mastering Bitcoin: unlocking digital cryptocurrencies*. " O'Reilly Media, Inc.".

arXiv (2017). arXiv monthly submission rates. https://arxiv.org/stats/monthly_submissions. Date accessed: 2017-06-12.

Baggs, J. G., Broome, M. E., Dougherty, M. C., Freda, M. C., and Kearney, M. H. (2008). Blinding in peer review: the preferences of reviewers for nursing journals. *Journal of advanced nursing*, 64(2):131–138, DOI: 10.1111/j.1365-2648.2008.04816.x.

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, DOI: 10.1098/rsnr.2015.0029.

Baldwin, M. (2017a). In referees we trust? *Physics Today*, 70(2):44–49, DOI: 10.1063/PT.3.3463.

Baldwin, M. (2017b). What it was like to be peer reviewed in the 1860s. *Physics Today*, DOI: 10.1063/PT.5.9098.

Ball, C. E., Lamanna, C. A., Saper, C., and Day, M. (2016). Annotated bibliography on evaluating digital scholarship for tenure and promotion. Stasis: A *Kairos* Archive, http://technorhetoric.net/stasis/2016/ball-et-al/Evaluating_Digital_Scholarship_for_Tenure.pdf.

Bartling, S. and Fecher, B. (2016). Blockchain for science and knowledge creation. DOI: 10.5281/zenodo.60223.

Baxt, W. G., Waeckerle, J. F., Berlin, J. A., and Callaham, M. L. (1998). Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. *Annals of Emergency Medicine*, 32(3):310–317, DOI: 10.1016/S0196-0644(98)70006-X.

Bedeian, A. G. (2003). The manuscript review process the proper roles of authors, referees, and editors. *Journal of Management Inquiry*, 12(4):331–338, DOI: 10.1177/1056492603258974.

Begel, A., Bosch, J., and Storey, M.-A. (2013). Social networking meets software development: Perspectives from GitHub, MSDN, Stack Exchange, and TopCoder. *IEEE Software*, 30(1):52–66, DOI: 10.1109/MS.2013.13.

Belojevic, N., Sayers, J., et al. (2014). Peer review personas. *Journal of Electronic Publishing*, 17(3), DOI: 10.3998/3336451.0017.304.

Benda, W. G. and Engels, T. C. (2011). The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science. *International Journal of Forecasting*, 27(1):166–182, DOI: 10.1016/j.ijforecast.2010.03.003.

Bernstein, R. (2015). Updated: Sexist peer review elicits furious twitter response, PLOS apology. *Science*, DOI: 10.1126/science.aab2568.

Berthaud, C., Capelli, L., Gustedt, J., Kirchner, C., Loiseau, K., Magron, A., Medves, M., Monteil, A., Rivrieux, G., and Romary, L. (2014). EPISCIENCES – an overlay publication platform. *Information Services Use*, 34(3-4):269–277, DOI: 10.3233/978-1-61499-409-1-78.

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures*, 12(1):11–45, DOI: 10.1080/1045722022000003435.

Black, E. W. (2008). Wikipedia and academic peer review: Wikipedia as a recognised medium for scholarly publication? *Online Information Review*, 32(1):73–88, DOI: 10.1108/14684520810865994.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review. *The American Economic Review*, 81(5):1041–1067, https://www.jstor.org/stable/2006906.

Boldt, A. (2011). Extending ArXiv.org to achieve open peer review and publishing. *Journal of Scholarly Publishing*, 42(2):238–242, DOI: 10.3138/jsp.42.2.238.

Bon, M., Taylor, M., and McDowell, G. S. (2017). Novel processes and metrics for a scientific evaluation rooted in the principles of science - Version 1. arXiv:1701.08008 [cs.DL].

Bornmann, L. and Daniel, H.-D. (2010). How long is the peer review process for journal manuscripts? A case study on *Angewandte Chemie International Edition*. *CHIMIA International Journal for Chemistry*, 64(1):72–77, DOI: 10.2533/chimia.2010.72.

Bornmann, L. and Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222, DOI: 10.1002/asi.23329.

Bornmann, L., Wolf, M., and Daniel, H.-D. (2012). Closed versus open reviewing of journal manuscripts: how far do comments differ in language use? *Scientometrics*, 91(3):843–856, DOI: 10.1007/s11192-011-0569-5.

Brembs, B. (2015). The cost of the rejection-resubmission cycle. *The Winnower*, DOI: 10.15200/winn.142497.72083.

Breuning, M., Backstrom, J., Brannon, J., Gross, B. I., and Widmeier, M. (2015). Reviewer fatigue? why scholars decline to review their peers' work. *PS: Political Science & Politics*, 48(4):595–600, DOI: 10.1017/S1049096515000827.

Budden, A., Tregenza, T., Aarssen, L., Koricheva, J., Leimu, R., and Lortie, C. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, 23(1):4–6, DOI: 10.1016/j.tree.2007.07.008.

Burghardt, K., Alsina, E. F., Girvan, M., Rand, W., and Lerman, K. (2017). The myopia of crowds: Cognitive load and collective evaluation of answers on stack exchange. *PLoS ONE*, 12(3):1–19, DOI: 10.1371/journal.pone.0173610.

Burnham, J. C. (1990). The evolution of editorial peer review. *JAMA*, 263(10):1323–1329, DOI: 10.1001/jama.1990.03440100023003.

Burris, V. (2004). The academic caste system: Prestige hierarchies in PhD exchange networks. *American Sociological Review*, 69(2):239–264, DOI: 10.1177/000312240406900205.

Campanario, J. M. (1998a). Peer review for journals as it stands today—part 1. *Science Communication*, 19(3):181–211, DOI: 10.1177/1075547098019003002.

Campanario, J. M. (1998b). Peer review for journals as it stands today—part 2. *Science Communication*, 19(4):277–306, DOI: 10.1177/1075547098019004002.

Carlisle, B. G. (2014). Proof of prespecified endpoints in medical research with the bitcoin blockchain. *The Grey Literature*, https://www.bgcarlisle.com/blog/2014/08/25/proof-of-prespecified-endpoints-in-medical-research-with-the-bitcoin-blockchain/. Date accessed: 2017-06-12.

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., and Etchells, P. J. (2014). Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1(1):4–17, DOI: 10.3934/Neuroscience.2014.1.4.

Chambers, C. D., Forstmann, B., and Pruszynski, J. A. (2017). Registered reports at the European Journal of Neuroscience: consolidating and extending peer-reviewed study pre-registration. *European Journal of Neuroscience*, 45(5):627–628, DOI: 10.1111/ejn.13519.

Chevalier, J. A. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, DOI: 10.1509/jmkr.43.3.345.

Cole, S. (2000). The role of journals in the growth of scientific knowledge. In Cronin, B. and Atkins, H. B., editors, *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, ASIS Monograph Series, chapter 6, pages 109–142. Information Today, Inc., Medford, NJ.

Cope, B. and Kalantzis, M. (2009). Signs of epistemic disruption: Transformations in the knowledge system of the academic journal. *The Future of the Academic Journal. Oxford: Chandos Publishing*, 14(2):13–61, DOI: 10.5210/fm.v14i4.2309.

Crotty, D. (2009). How meaningful are user ratings? (this article = 4.5 stars!). *The Scholarly Kitchen*, https://scholarlykitchen.sspnet.org/2009/11/16/how-meaningful-are-user-ratings-this-article-4-5-stars/. Date accessed: 2017-06-12.

Dall'Aglio, P. (2006). Peer review and journal models. arXiv:physics/0608307 [physics.soc-ph].

D'Andrea, R. and O'Dwyer, J. P. (2017). Can editors protect peer review from bad reviewers? *PeerJ Preprints*, 5:e3005v1, DOI: 10.7287/peerj.preprints.3005v1.

Dappert, A., Farquhar, A., Kotarski, R., and Hewlett, K. (2017). Connecting the persistent identifier ecosystem: Building the technical and human infrastructure for open research. *Data Science Journal*, 16.

Darling, E. (2015). Use of double-blind peer review to increase author diversity. *Conservation Biology*, 29(1):297–299, DOI: 10.1111/cobi.12333.

Davis, P. (2017). Wither portable peer review. *The Scholarly Kitchen*, https://scholarlykitchen.sspnet.org/2017/03/20/wither-portable-peer-review. Accessed: 2017-06-12.

Davis, P. and Fromerth, M. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2):203–215, DOI: 10.1007/s11192-007-1661-8.

Dhillon, V. (2016). From bench to bedside: Enabling reproducible commercial science via blockchain. *Bitcoin Magazine*, https://bitcoinmagazine.com/

articles/from-bench-to-bedside-enabling-reproducible-commercial-science-via-blockchain-1464881141/. Date accessed: 2017-06-12.

Drummond, R. (2016). Let's make peer review scientific. *Nature*, 535:31–33, DOI: 10.1038/535031a.

Eckberg, D. L. (1991). When nonreliability of reviews indicates solid science. *Behavioral and Brain Sciences*, 14(01):145–146, DOI: 10.1017/S0140525X00065791.

Edgar, B. and Willinsky, J. (2010). A survey of scholarly journals using open journal systems. *Scholarly and Research Communication*, 1(2), DOI: 10.22230/src.2010v1n2a24.

Eisen, M. (2011). Peer review is f***ed up – let's fix it. http://www.michaeleisen.org/blog/?p=694. Accessed: 2017-03-15.

Elkhatib, Y., Tyson, G., and Sathiaseelan, A. (2015). Does the Internet deserve everybody? In *Proceedings of the 2015 ACM SIGCOMM Workshop on Ethics in Networked Systems Research*, pages 5–8. ACM, ISBN: 978-1-4503-3541-6, DOI: 10.1145/2793013.2793018.

Emilsson, R. (2015). The influence of the Internet on identity creation and extreme groups. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A838907&dswid=7386. Bachelor Thesis, Blekinge Institute of Technology, Department of Technology and Aesthetics.

Ernst, E. and Kienbacher, T. (1991). Chauvinism. *Nature*, 352:560, DOI: 10.1038/352560b0.

Farley, T. (2011). Hypothes.is reaches funding goal. *James Randi Educational Foundation Swift Blog*, http://archive.randi.org/site/index.php/swift-blog/1524-hypothesis-reaches-funding-goal.html. Date accessed: 2017-06-12.

Fitzpatrick, K. (2010). Peer-to-peer review and the future of scholarly authority. *Social Epistemology*, 24(3):161–179, DOI: 10.1080/02691728.2010.498929.

Fitzpatrick, K. (2011a). Peer review, judgment, and reading. *Profession*, pages 196–201, DOI: 10.1632/prof.2011.2011.1.196.

Fitzpatrick, K. (2011b). *Planned Obsolescence*. New York University Press, New York, ISBN: 978-0-8147-2788-1 0-8147-2788-3, http://mcpress.media-commons.org/plannedobsolescence/.

Ford, E. (2013). Defining and characterizing open peer review: A review of the literature. *Journal of Scholarly Publishing*, 44(4):311–326, DOI: 10.3138/jsp.44-4-001.

Fox, C. W., Albert, A. Y., and Vines, T. H. (2017). Recruitment of reviewers is becoming harder at some journals: a test of the influence of reviewer fatigue at six journals in ecology and evolution. *Research Integrity and Peer Review*, 2(1):3, DOI: 10.1186/s41073-017-0027-x.

Fox, M. F. (1994). Scientific misconduct and editorial and peer review processes. *The Journal of Higher Education*, 65(3):298–309.

Frishauf, P. (2009). Reputation systems: a new vision for publishing and peer review. *Journal of Participatory Medicine*, 1(1):e13a, http://www.jopm.org/opinion/commentary/2009/10/21/reputation-systems-a-new-vision-for-publishing-and-peer-review/.

Fukuzawa, N. (2017). Characteristics of papers published in journals: an analysis of open access journals, country of publication, and languages used. *Scientometrics*, pages 1–17, ISSN: 1588-2861, DOI: 10.1007/s11192-017-2414-y, http://dx.doi.org/10.1007/s11192-017-2414-y.

Fyfe, A. (2015). Journals, learned societies and money: *Philosophical Transactions*, ca. 1750–1900. *Notes and Records: the Royal Society Journal of the History of Science*, 69(3):277–299, DOI: 10.1098/rsnr.2015.0032.

Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N., and Røstvik, C. M. (2017). Untangling Academic Publishing: A history of the relationship between commercial interests, academic prestige and the circulation of research. Zenodo, DOI: 10.5281/zenodo.546100.

Gashler, M. (2008). GPeerReview - a tool for making digital-signatures using data mining. *KDnuggets*, http://kdnuggets.com/news/2008/n16/5i.html. Date accessed: 2017-06-12.

Ghosh, S., Klein, A., Avants, B., and Millman, K. (2012). Learning from open source software projects to improve scientific review. *Frontiers in Computational Neuroscience*, 6:18, DOI: 10.3389/fncom.2012.00018.

Gibney, E. (2016). Toolbox: Low-cost journals piggyback on arXiv. *Nature*, 530(7588):117–118, DOI: 10.1038/530117a.

Gibson, M., Spong, C. Y., Simonsen, S. E., Martin, S., and Scott, J. R. (2008). Author perception of peer review. *Obstetrics & Gynecology*, 112(3):646–652, DOI: 10.1097/AOG.0b013e31818425d4.

Ginsparg, P. (1997). Winners and losers in the global research village. *The Serials Librarian*, 30(3–4):83–95, DOI: 10.1300/J123v30n03_13.

Godlee, F. (2002). Making reviewers visible: Openness, accountability, and credit. *JAMA*, 287(21):2762–2765, DOI: 10.1001/jama.287.21.2762.

Godlee, F., Gale, C., and Martyn, C. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *JAMA*, 280(3):237–240, DOI: 10.1001/jama.280.3.237.

Goodman, S. N., Berlin, J., Fletcher, S. W., and Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at annals of internal medicine. *Annals of Internal Medicine*, 121(1):11–21, DOI: 10.7326/0003-4819-121-1-199407010-00003.

Gøtzsche, P. C. (1989). Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clinical Trials*, 10(1):31–56, DOI: 10.1016/0197-2456(89)90017-2.

Graf, K. (2014). Fetisch peer review. Archivalia, http://archivalia.hypotheses.org/4285. Accessed: 2017-03-15; in German.

Greaves, S., Scott, J., Clarke, M., Miller, L., Hannay, T., Thomas, A., and Campbell, P. (2006). Overview: Nature's peer review trial. *Nature*, DOI: 10.1038/nature05535.

Greenberg, S. A. (2009). How citation distortions create unfounded authority: analysis of a citation network. *BMJ*, 339:b2680.

Grivell, L. (2006). Through a glass darkly. *EMBO reports*, 7(6):567–570, DOI: 10.1038/sj.embor.7400718.

Gropp, R. E., Glisson, S., Gallo, S., and Thompson, L. (2017). Peer review: A system under stress. *BioScience*, 67(5):407–410, DOI: 10.1093/biosci/bix034.

Gupta, S. (2016). How has publishing changed in the last twenty years? *Notes and Records: the Royal Society Journal of the History of Science*, 70(4):391–392, DOI: 10.1098/rsnr.2016.0035.

Haider, J. and Åström, F. (2017). Dimensions of trust in scholarly communication: Problematizing peer review in the aftermath of John Bohannon's "sting" in science. *Journal of the Association for Information Science and Technology*, 68(2):450–467, DOI: 10.1002/asi.23669.

Halavais, A., Kwon, K. H., Havener, S., and Striker, J. (2014). Badges of friendship: Social influence and badge acquisition on stack overflow. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 1607–1615. IEEE, DOI: 10.1109/HICSS.2014.206.

Harley, D., Lawrence, S., Acord, S. K., and Dixson, J. (2010). Affordable and open textbooks: An exploratory study of faculty attitudes. *The California Journal of Politics and Policy*, 2, DOI: 10.5070/P2D60T.

Harmon, A. H. and Metaxas, P. T. (2010). How to create a smart mob: Understanding a social network capital. In Krishnamurthy, S., Singh, G., and McPherson, M., editors, *Proceedings of the IADIS International Conference on e-Democracy, Equity and Social Justice*. International Association for Development of the Information Society, ISBN: 978-972-8939-24-3.

Hasty, R., Garvalosa, R., Barbato, V, Valdes, P, Powers, D., Hernandez, E., John, J., Suciu, G., Qureshi, F., Popa-Radu, M., Jose, S. S., Drexler, N., Patankar, R., Paz, J., King, C., Gerber, H., Valladares, M., and Somji, A. (2014). Wikipedia vs peer-reviewed medical literature for information about the 10 most costly medical conditions. *The Journal of the American Osteopathic Association*, 114(5):368–373, DOI: 10.7556/jaoa.2014.035.

Haug, C. J. (2015). Peer-review fraud — hacking the scientific publication process. *New England Journal of Medicine*, 373(25):2393–2395, DOI: 10.1056/NEJMp1512330.

Heaberlin, B. and DeDeo, S. (2016). The evolution of wikipedia's norm network. *Future Internet*, 8(2):14.

Heller, L., The, R., and Bartling, S. (2014). *Dynamic Publication Formats and Collaborative Authoring*, pages 191–211. Springer International Publishing, Cham, ISBN: 978-3-319-00026-8, DOI: 10.1007/978-3-319-00026-8_13.

Helmer, M., Schottdorf, M., Neef, A., and Battaglia, D. (2017). Gender bias in scholarly peer review. *eLife*, 6:e21718, DOI: 10.7554/eLife.21718.

Hettyey, A., Griggio, M., Mann, M., Raveh, S., Schaedelin, F. C., Thonhauser, K. E., Thoß, M., van Dongen, W. F., White, J., Zala, S. M., and Penn, D. J. (2012). Peerage of Science: will it work? *Trends in Ecology & Evolution*, 27(4):189–190, DOI: 10.1016/j.tree.2012.01.005.

Horrobin, D. F. (1990). The philosophical basis of peer review and the suppression of innovation. *JAMA*, 263(10):1438–1441, DOI: 10.1001/jama.1990.03440100162024.

Hu, M., Lim, E.-P., Sun, A., Lauw, H. W., and Vuong, B.-Q. (2007). Measuring article quality in Wikipedia: models and evaluation. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pages 243–252. ACM, DOI: 10.1145/1321440.1321476.

Hukkinen, J. I. (2017). Peer review has its shortcomings, but AI is a risky fix. *Wired*, https://www.wired.com/2017/01/peer-review-shortcomings-ai-risky-fix/. Accessed: 2017-01-31.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2(8):e124, DOI: 10.1371/journal.pmed.0020124.

Isenberg, S. J., Sanchez, E., and Zafran, K. C. (2009). The effect of masking manuscripts for the peer-review process of an ophthalmic journal. *British Journal of Ophthalmology*, 93(7):881–884, DOI: 10.1136/bjo.2008.151886.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., and McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled clinical trials*, 17(1):1–12, DOI: 10.1016/0197-2456(95)00134-4.

Janowicz, K. and Hitzler, P. (2012). Open and transparent: the review process of the semantic web journal. *Learned Publishing*, 25(1):48–55, DOI: 10.1087/20120107.

Jefferson, T., Wager, E., and Davidoff, F. (2002). Measuring the quality of editorial peer review. *Jama*, 287(21):2786–2790, DOI: 10.1001/jama.287.21.2786.

Jubb, M. (2016). Peer review: The current landscape and future trends. *Learned Publishing*, 29(1):13–21.

Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., Rennie, D., Investigators, P., et al. (1998). Does masking author identity improve peer review quality?: A randomized controlled trial. *Journal of the American Medical Association*, 280(3):240–242, DOI: 10.1001/jama.280.3.240.

Katsh, E. and Rule, C. (2015). What we know and need to know about online dispute resolution. *SCL Rev.*, 67:329.

Kelty, C. M., Burrus, C. S., and Baraniuk, R. G. (2008). Peer review anew: Three principles and a case study in postpublication quality assurance. *Proceedings of the IEEE*, 96(6):1000–1011.

Khan, M. S. (2012). Exploring citations for conflict of interest detection in peer review system. *International Journal of Computer Information Systems and Industrial Management Applications*, 4:283–299.

Klyne, G. (2016). Peer review #2 of "software citation principles (v0.1)". *PeerJ Computer Science*, DOI: 10.7287/peerj-cs.86v0.1/reviews/2.

Kosner, A. W. (2012). GitHub is the next big social network, powered by what you do, not who you know. *Forbes*, https://www.forbes.com/sites/anthonykosner/2012/07/15/github-is-the-next-big-social-network-powered-by-what-you-do-not-who-you-know/#483d992a45ce.

Kostoff, R. and N, Q. (1995). Federal research impact assessment: Axioms, approaches, applications. *Scientometrics*, 34(2):163–206, DOI: 10.1007/BF02020420.

Kovanis, M., Porcher, R., Ravaud, P, and 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, DOI: 10.1371/journal.pone.0166387.

Kriegeskorte, N., Walther, A., and Deca, D. (2012). An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. *Frontiers in Computational Neuroscience*, 6:94, ISSN: 1662-5188, DOI: 10.3389/fncom.2012.00094, http://journal.frontiersin.org/article/10.3389/fncom.2012.00094.

Kronick, D. A. (1990). Peer review in 18th-century scientific journalism. *JAMA*, 263(10):1321–1322, DOI: 10.1001/jama.1990.03440100021002.

Kubátová, J. et al. (2012). Growth of collective intelligence by linking knowledge workers through social media. *Lex ET Scientia International Journal (LESIJ)*, (XIX-1):135–145.

Kuhn, T. (2016a). Peer review #1 of "software citation principles (v0.1)". *PeerJ Computer Science*, DOI: 10.7287/peerj-cs.86v0.1/reviews/1.

Kuhn, T. (2016b). Peer review #1 of "software citation principles (v0.2)". *PeerJ Computer Science*, DOI: 10.7287/peerj-cs.86v0.2/reviews/1.

Larivière, V, Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., and Thelwall, M. (2014). arxiv e-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology*, 65(6):1157–1169, DOI: 10.1002/asi.23044.

Larivière, V, Haustein, S., and Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE*, 10(6):1–15, DOI: 10.1371/journal.pone.0127502, https://doi.org/10.1371/journal.pone.0127502.

Larsen, P. O. and Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by science citation index. *Scientometrics*, 84(3):575–603, DOI: 10.1007/s11192-010-0202-z.

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1):2–17, DOI: 10.1002/asi.22784.

Lee, D. (2015). The new Reddit journal of science. *IMMpress Magazine*, http://www.immpressmagazine.com/the-new-reddit-journal-of-science/. Date accessed: 2016-06-12.

Leek, J. T., Taub, M. A., and Pineda, F. J. (2011). Cooperation between referees and authors increases peer review accuracy. *PLOS ONE*, 6(11):e26895, DOI: 10.1371/journal.pone.0026895.

Lerback, J. and Hanson, B. (2017). Journals invite too few women to referee. *Nature*, 541:455–457, DOI: 10.1038/541455a.

Li, L., Steckelberg, A., and Srinivasan, S. (2009). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology*, 34(2), DOI: 10.21432/T21C7R.

Link, A. (1998). US and non-US submissions: An analysis of reviewer bias. *JAMA*, 280(3):246–247, DOI: 10.1001/jama.280.3.246.

Lipworth, W., Kerridge, I. H., Carter, S. M., and Little, M. (2011). Should biomedical publishing be "opened up"? toward a values-based peer-review process. *Journal of Bioethical Inquiry*, 8(3):267–280, DOI: 10.1007/s11673-011-9312-4.

List, B. (2017). Crowd-based peer review can be good and fast. *Nature*, 546(7656):9, DOI: 10.1038/546009a.

Lloyd, M. E. (1990). Gender factors in reviewer recommendations for manuscript publication. *Journal of Applied Behavior Analysis*, 23(4):539–543, DOI: 10.1901/jaba.1990.23-539.

Lui, K. M. and Chan, K. C. (2006). Pair programming productivity: Novice–novice vs. expert–expert. *International Journal of Human-computer studies*, 64(9):915–925, DOI: 10.1016/j.ijhcs.2006.04.010.

Luzi, D. (2000). Trends and evolution in the development of grey literature: a review. *International Journal on Grey Literature*, 1(3):106–117, DOI: 10.1108/14666180010345537.

Lyman, R. L. (2013). A three-decade history of the duration of peer review. *Journal of Scholarly Publishing*, 44(3):211–220, DOI: 10.3138/jsp.44.3.001.

Magee, J. C. and Galinsky, A. D. (2008). 8 social hierarchy: The self-reinforcing nature of power and status. *The Academy of Management Annals*, 2(1):351–398, DOI: 10.1080/19416520802211628.

Maharg, P. and Duncan, N. (2007). Black box, pandora's box or virtual toolbox? an experiment in a journal's transparent peer review on the web. *International Review of Law Computers & Technology*, 21(2):109–128, DOI: 10.1080/13600860701492104.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2):161–175, DOI: 10.1007/BF01173636.

Manten, A. A. (1980). Development of european scientific journal publishing before 1850. *Development of science publishing in Europe*, pages 1–22.

Margalida, A. and Colomer, M. A. (2016). Improving the peer-review process and editorial quality: key errors escaping the review and editorial process in top scientific journals. *PeerJ*, page e1670, DOI: 10.7717/peerj.1670.

Marra, M. (2017). Arxiv-based commenting resources by and for astrophysicists and physicists: An initial survey. *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, pages 100–117, DOI: 10.3233/978-1-61499-769-6-100.

McCormack, N. (2009). Peer review and legal publishing: What law librarians need to know about open, single-blind, and double-blind reviewing. *Law Library Journal*, 101:59.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., and Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5:e16800, DOI: 10.7554/eLife.16800.

McKiernan, G. (2003). Alternative peer review: Quality management for 21st century scholarship. Talk presented at the Workshop on Peer Review in the Age of Open Archives, in Trieste, Italy, http://www.public.iastate.edu/~gerrymck/APR-1.ppt.

McNutt, R. A., Evans, A. T., Fletcher, R. H., and Fletcher, S. W. (1990). The effects of blinding on the quality of peer review: a randomized trial. *Journal of the American Medical Association*, 263(10):1371–1376, DOI: 10.1001/jama.1990.03440100079012.

Melero, R. and Lopez-Santovena, F. (2001). Referees' attitudes toward open peer review and electronic transmission of papers. *Food Science and Technology International*, 7(6):521–527, DOI: 10.1106/0MXD-YM6F-3LM6-G9EB.

Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810):56–63, DOI: 10.1126/science.159.3810.56.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press, Chicago.

Mhurchú, A. N., McLeod, L., Collins, S., and Siles-Brügge, G. (2017). The present and the future of the research excellence framework impact agenda in the UK academy: A reflection from politics and international studies. *Political Studies Review*, 15(1):60–72, DOI: 10.1177/1478929916658918.

Moed, H. F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology*, 58(13):2047–2054, DOI: 10.1002/asi.20663.

Moore, S., Neylon, C., Eve, M. P., O'Donnell, D. P., and Pattinson, D. (2017). "excellence R Us": university research and the fetishisation of excellence. *Palgrave Communications*, 3:16105, DOI: 10.1057/palcomms.2016.105.

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C. C., Newman, D. P., Schönbrodt, F. D., Vanpaemel, W., Wagenmakers, E.-J., and Zwaan, R. A. (2016). The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *Royal Society Open Science*, 3(1), DOI: 10.1098/rsos.150547.

Morrison, J. (2006). The case for open peer review. *Medical Education*, 40(9):830–831, DOI: 10.1111/j.1365-2929.2006.02573.x.

Moxham, N. and Fyfe, A. (2016). A pre-history of 'peer review': refereeing and editorial selection at the royal society. *Historical Journal*, http://hdl.handle.net/10023/9434.

Mudambi, S. M. and Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34:185–200, https://ssrn.com/abstract=2175066.

Mulligan, A., Akerman, R., Granier, B., Tamber, P. S., and Pöschl, U. (2008). Quality, certification and peer review. *Information Services and Use*, 28(3-4):197–214, http://dl.acm.org/citation.cfm?id=1498587.1498596.

Mulligan, A., Hall, L., and Raphael, E. (2013). Peer review in a changing world: An international study measuring the attitudes of researchers. *Journal of the American Society for Information Science and Technology*, 64(1):132–161, DOI: 10.1002/asi.22798.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., and Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1:0021.

Murphy, T. and Sage, D. (2014). Perceptions of the UK's research excellence framework 2014: a media analysis. *Journal of Higher Education Policy and Management*, 36(6):603–615, DOI: 10.1080/1360080X.2014.957890.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Nature (2010). Response required. *Nature*, 468(7326):867–867, DOI: 10.1038/468867a.

Nature Human Behaviour (2017). Promoting reproducibility with registered reports. *Nature Human Behaviour*, 1:0034, DOI: 10.1038/s41562-016-0034.

Neylon, C. and Wu, S. (2009). Article-level metrics and the evolution of scientific impact. *PLoS Biol*, 7(11):e1000242, DOI: 10.1371/journal.pbio.1000242.

Nicholson, J. and Alperin, J. P. (2016). A brief survey on peer review in scholarly communication. *The Winnower*, https://thewinnower.com/papers/4659-a-brief-survey-on-peer-review-in-scholarly-communication.

Nobarany, S. and Booth, K. S. (2016). Understanding and supporting anonymity policies in peer review. *Journal of the Association for Information Science and Technology*, 68(4):957–971, DOI: 10.1002/asi.23711.

Nosek, B. A. and Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3):137–141, DOI: 10.1027/1864-9335/a000192.

Okike, K., Hug, K. T., Kocher, M. S., and Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *JAMA*, 316(12):1315, DOI: 10.1001/jama.2016.11014.

Oldenburg, H. (1665). Epistle dedicatory. *Philosophical Transactions*, 1(1–22), DOI: 10.1098/rstl.1665.0001.

Open Access Textbook Task Force (2010). Open Access Textbook Task Force Final Report. Technical report, State of Florida, https://florida.theorangegrove.org/og/items/a83149bc-c972-ed56-67ae-b2952115ef08/1/.

Owens, S. (2014). The world's largest 2-way dialogue between scientists and the public. *Scientific American*, https://scientificamerican.com/article/the-world-s-largest-2-way-dialogue-between-scientists-and-the-public. Date accessed: 2017-06-12.

Paglione, L. D. and Lawrence, R. N. (2015). Data exchange standards to support and acknowledge peer-review activity. *Learned Publishing*, 28(4):309–316, DOI: 10.1087/20150411.

Pallavi Sudhir, A. and Knöpfel, R. (2015). PhysicsOverflow: A postgraduate-level physics Q&A site and open peer review system. *Asia Pacific Physics Newsletter*, 4(1):53–55, DOI: 10.1142/S2251158X15000193.

Parnell, L. D., Lindenbaum, P, Shameer, K., Dall'Olio, G. M., Swan, D. C., Jensen, L. J., Cockell, S. J., Pedersen, B. S., Mangan, M. E., Miller, C. A., and Albert, I. (2011). BioStar: An online question & answer resource for the bioinformatics community. *PLOS Computational Biology*, 7(10):1–5, DOI: 10.1371/journal.pcbi.1002216.

Patel, J. (2014). Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *BMC Medicine*, 12:128, DOI: 10.1186/s12916-014-0128-z.

Perakakis, P, Taylor, M., Mazza, M., and Trachana, V. (2010). Natural selection of academic papers. *Scientometrics*, 85(2):553–559, DOI: 10.1007/s11192-010-0253-1.

Perkel, J. M. (2015). Annotating the scholarly web. *Nature*, 528(7580):153, DOI: 10.1038/528153a.

Peters, D. P. and Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2):187–195, DOI: 10.1017/S0140525X00011183.

Petrides, L., Jimes, C., Middleton-Detzner, C., Walling, J., and Weiss, S. (2011). Open textbook adoption and use: implications for teachers and learners. *Open Learning: The Journal of Open, Distance and e-Learning*, 26(1):39–49, DOI: 10.1080/02680513.2011.538563.

Pierie, J.-P. E., Walvoort, H. C., and Overbeke, A. J. P. (1996). Readers' evaluation of effect of peer review and editing on quality of articles in the nederlands tijdschrift voor geneeskunde. *The Lancet*, 348(9040):1480–1483, DOI: 10.1016/S0140-6736(96)05016-7.

Pinfield, S. (2016). Mega-journals: the future, a stepping stone to it or a leap into the abyss? *Times Higher Education*, https://timeshighereducation.com/blog/mega-journals-future-stepping-stone-it-or-leap-abyss. Date accessed: 2017-06-12.

Plume, A. and van Weijen, D. (2014). Publish or perish? The rise of the fractional author. . . . *Research Trends*, 38(3).

Pocock, S. J., Hughes, M. D., and Lee, R. J. (1987). Statistical problems in the reporting of clinical trials. *New England Journal of Medicine*, 317(7):426–432, DOI: 10.1056/NEJM198708133170706.

Pontille, D. and Torny, D. (2014). The blind shall see! the question of anonymity in journal peer review. *Ada: A Journal of Gender, New Media, and Technology*, 4, DOI: 10.7264/N3542KVW.

Pontille, D. and Torny, D. (2015). From manuscript evaluation to article valuation: the changing technologies of journal peer review. *Human Studies*, 38(1):57–79, DOI: 10.1007/s10746-014-9335-z.

Prechelt, L., Graziotin, D., and Fernández, D. M. (2017). On the status and future of peer review in software engineering. *arXiv*, DOI: arXiv:1706.07196 [cs.SE].

Priem, J. (2013). Scholarship: Beyond the paper. *Nature*, 495(7442):437–440, DOI: 10.1038/495437a.

Priem, J. and Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social web. *First Monday*, 15(7), DOI: 10.5210/fm.v15i7.2874.

Priem, J. and Hemminger, B. M. (2012). Decoupling the scholarly journal. *Frontiers in Computational Neuroscience*, 6:19, DOI: 10.3389/fncom.2012.00019.

Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., and Asgari-Targhi, M. (2010a). Adoption and use of web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1926):4039–4056.

Procter, R. N., Williams, R., Stewart, J., et al. (2010b). If you build it, will they come? How researchers perceive and use Web 2.0. Technical report, Research Network Information, http://www.rin.ac.uk/system/files/attachments/web_2.0_screen.pdf.

Public Knowledge Project (2016). OJS stats. https://pkp.sfu.ca/ojs/ojs-usage/ojs-stats. Date accessed: 2017-06-12.

Pullum, G. K. (1984). Stalking the perfect journal. *Natural language & linguistic theory*, 2(2):261–267, ISSN: 0167-806X, http://www.jstor.org/stable/4047489.

Rennie, D. (2003). Misconduct and journal peer review.

Research Information Network (2008). Activities, costs and funding flows in the scholarly communications system in the UK: Report commissioned by the Research Information Network (RIN). http://www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf.

Review Meta (2016). Analysis of 7 million Amazon reviews: customers who receive free or discounted item much more likely to write positive review. https://reviewmeta.com/blog/analysis-of-7-million-amazon-reviews-customers-who-receive-free-or-discounted-item-much-more-likely-to-write-positive-review/. Date accessed: 2017-06-12.

Riggs, J. E. (1995). Priority, rivalry, and peer review. *Journal of Child Neurology*, 10(3):255–256, DOI: 10.1177/088307389501000325.

Roberts, S. G. and Verhoef, T. (2016). Double-blind reviewing at evolang 11 reveals gender bias. *Journal of Language Evolution*, 1(2):163–167, DOI: 10.1093/jole/lzw009.

Rodriguez, M. A. and Bollen, J. (2008). An algorithm to determine peer-reviewers. In *Proceedings of the 17th ACM conference on Information and Knowledge Management*, pages 319–328. ACM, DOI: 10.1145/1458082.1458127.

Rodriguez, M. A., Bollen, J., and Van de Sompel, H. (2006). The convergence of digital libraries and the peer-review process. *Journal of Information Science*, 32(2):149–159, DOI: 10.1177/0165551506062327.

Ross, J., Gross, C., Desai, M., and et al (2006). Effect of blinded peer review on abstract acceptance. *JAMA*, 295(14):1675–1680, DOI: 10.1001/jama.295.14.1675.

Ross, N., Boettiger, C., Bryan, J., Chamberlain, S., FitzJohn, R., and Ram, K. (2016). Onboarding at rOpenSci: A year in reviews. *rOpenSci Blog*, https://ropensci.org/blog/2016/03/28/software-review. Date accessed: 2017-06-12.

Ross-Hellauer, T. (2017). What is open peer review? a systematic review [version 1; referees: awaiting peer review]. *F1000Research*, 6(588), DOI: 10.12688/f1000research.11369.1.

Rughiniş, R. and Matei, S. (2013). Digital badges: Signposts and claims of achievement. In Stephanidis, C., editor, *HCI International 2013 - Posters' Extended Abstracts. HCI 2013. Communications in Computer and Information Science*, volume 374, pages 84–88. Springer, Berlin, Heidelberg, ISBN: 978-3-642-39476-8, DOI: 10.1007/978-3-642-39476-8_18.

Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for the future. *Journal of English for Academic Purposes*, 7(2):121–132.

Salager-Meyer, F. (2014). Writing and publishing in peripheral scholarly journals: How to enhance the global influence of multilingual scholars? *Journal of English for Academic Purposes*, 13:78–82.

Sanger, L. (2005). The early history of Nupedia and Wikipedia: a memoir. In DiBona, C., Stone, M., and Cooper, D., editors, *Open Sources 2.0: The Continuing Evolution*, chapter 20, pages 307–338. O'Reilly Media, Sebastopol, CA, ISBN: 978-0-596-00802-4.

Schiermeier, Q. (2017). 'you never said my peer review was confidential' — scientist challenges publisher. *Nature*, 541:446, DOI: 10.1038/nature.2017.21342.

Schmidt, B. and Görögh, E. (2017). New toolkits on the block: Peer review alternatives in scholarly communication. In Chan, L. and Loizides, F., editors, *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, pages 62–74. IOS Press, DOI: 10.3233/978-1-61499-769-6-62.

Schroter, S., Black, N., Evans, S., Carpenter, J., Godlee, F., and Smith, R. (2004). Effects of training on quality of peer review: randomised controlled trial. *BMJ*, 328(7441):673, DOI: 10.1136/bmj.38023.700775.AE.

Schroter, S., Tite, L., Hutchings, A., and Black, N. (2006). Differences in review quality and recommendations for publication between peer reviewers suggested by authors or by editors. *Journal of the American Medical Association*, 295(3):314–317, DOI: 10.1001/jama.295.3.314.

Shotton, D. (2012). The five stars of online journal articles: A framework for article evaluation. *D-Lib Magazine*, 18(1):1, DOI: 10.1045/january2012-shotton.

Shuttleworth, S. and Charnley, B. (2016). Science periodicals in the nineteenth and twenty-first centuries. *Notes and Records: the Royal Society Journal of the History of Science*, 70(4):297–304, DOI: 10.1098/rsnr.2016.0026.

Siler, K., Lee, K., and Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, 112(2):360–365, DOI: 10.1073/pnas.1418218112.

Singh Chawla, D. (2016). Here's why more than 50,000 psychology studies are about to have PubPeer entries. *Retraction Watch*, http://retractionwatch.com/2016/09/02/heres-why-more-than-50000-psychology-studies-are-about-to-have-pubpeer-entries. Date accessed: 2017-06-12.

Smith, A. M., Katz, D. S., Niemeyer, K. E., and FORCE11 Software Citation Working Group (2016). Software citation principles. *PeerJ Computer Science*, 2:e86, DOI: 10.7717/peerj-cs.86.

Smith, J. W. T. (1999). The deconstructed journal — a new model for academic publishing. *Learned Publishing*, 12(2):79–91, DOI: 10.1087/09531519950145896.

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4):178–182, DOI: 10.1258/jrsm.99.4.178.

Smith, R. (2010). Classical peer review: an empty gun. *Breast Cancer Research*, 12(4):S13, DOI: 10.1186/bcr2742.

Snell, L. and Spencer, J. (2005). Reviewers' perceptions of the peer review process for a medical education journal. *Medical Education*, 39(1):90–97, DOI: 10.1111/j.1365-2929.2004.02026.x.

Snodgrass, R. T. (2007). Editorial: Single-versus double-blind reviewing. *ACM Transactions on Database Systems*, 32(1):1, DOI: 10.1145/1206049.1206050.

Sobkowicz, P. (2008). Peer-review in the internet age. arXiv:0810.0486 [physics.soc-ph].

Spier, R. (2002). The history of the peer-review process. *TRENDS in Biotechnology*, 20(8):357–358, DOI: 10.1016/S0167-7799(02)01985-6.

Squazzoni, F., Grimaldo, F., and Marušić, A. (2017). Publishing: Journals should share peer review data. *Nature*, 546, DOI: 10.1038/546352a.

Stanton, D. C., Bérubé, M., Cassuto, L., Eaves, M., Guillory, J., Hall, D. E., and Latham, S. (2007). Report of the MLA task force on evaluating scholarship for tenure and promotion. *Profession*, pages 9–71, DOI: 10.1632/prof.2007.2007.1.9.

Steen, R. G., Casadevall, A., and Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLOS ONE*, 8(7):e68397, DOI: 10.1371/journal.pone.0068397.

Stemmle, L. and Collier, K. (2013). RUBRIQ: tools, services, and software to improve peer review. *Learned Publishing*, 26(4):265–268, DOI: 10.1087/20130406.

Swan, M. (2015). *Blockchain: Blueprint for a new economy*. O'Reilly Media, Sebastopol, CA, ISBN: 978-1-4919-2044-2.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. In *International Conference on Learning Representations*. Available via arXiv:1312.6199 [cs.CV].

Tausczik, Y. R., Kittur, A., and Kraut, R. E. (2014). Collaborative problem solving: A study of MathOverflow. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*, pages 355–367. ACM, DOI: 10.1145/2531602.2531690.

Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., and Hartgerink, C. H. (2016). The academic, economic and societal impacts of open access: an evidence-based review. *F1000Research*, 5, DOI: 10.12688/f1000research.8460.3.

Teytelman, L., Stoliartchouk, A., Kindler, L., and Hurwitz, B. L. (2016). Protocols.io: Virtual communities for protocol development and discussion. *PLoS Biology*, 14(8):e1002538, DOI: 10.1371/journal.pbio.1002538.

Thung, F., Bissyande, T. F., Lo, D., and Jiang, L. (2013). Network structure of social coding in GitHub. In *17th European Conference on Software Maintenance and Reengineering (CSMR)*, pages 323–326. IEEE, DOI: 10.1109/CSMR.2013.41.

Tomkins, A., Zhang, M., and Heavlin, W. D. (2017). Single versus double blind reviewing at WSDM 2017. arXiv:1702.00502 [cs.DL].

Torpey, K. (2015). Astroblocks puts proofs of scientific discoveries on the bitcoin blockchain. *Inside Bitcoins*, http://insidebitcoins.com/news/astroblocks-puts-proofs-of-scientific-discoveries-on-the-bitcoin-blockchain/31153. Date accessed: 2017-06-12.

Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology & Evolution*, 17(8):349–350, DOI: 10.1016/S0169-5347(02)02545-4.

Ubois, J. (2003). Online reputation systems. In Dyson, E., editor, *Release 1.0*, volume 21, pages 1–35. EDventure Holdings Inc., New York, NY, http://cdn.oreillystatic.com/radar/r1/10-03.pdf.

van Assen, M. A. L. M., van Aert, R. C. M., Nuijten, M. B., and Wicherts, J. M. (2014). Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS ONE*, 9(1):e84896, DOI: 10.1371/journal.pone.0084896.

Van Noorden, R. (2017). Web of Science owner buys up booming peer-review platform. *Nature News*, DOI: 10.1038/nature.2017.22094.

van Rooyen, S., Delamothe, T., and Evans, S. J. (2010). Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. *BMJ*, 341:c5729, DOI: 10.1136/bmj.c5729.

Van Rooyen, S., Godlee, F., Evans, S., Black, N., and Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ*, 318(7175):23–27, DOI: 10.1136/bmj.318.7175.23.

van Rooyen, S., Godlee, F., Evans, S., Smith, R., and Black, N. (1998). Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*, 280(3):234–237, DOI: 10.1001/jama.280.3.234.

Vines, T. (2015a). Molecular Ecology's best reviewers 2015. *The Molecular Ecologist*, http://www.molecularecologist.com/2015/04/mol-ecols-best-reviewers-2015/. Accessed: 2017-06-12.

Vines, T. H. (2015b). The core inefficiency of peer review and a potential solution. *Limnology and Oceanography Bulletin*, 24(2):36–38, DOI: 10.1002/lob.10022.

Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C. J., and Buytaert, W. (2015). Web technologies for environmental big data. *Environmental Modelling & Software*, 63:185 – 198, ISSN: 1364-8152, DOI: http://dx.doi.org/10.1016/j.envsoft.2014.10.007.

von Muhlen, M. (2011). We need a Github of science. http://marciovm.com/i-want-a-github-of-science. Date accessed: 2017-06-12.

W3C (2017). Three recommendations to enable annotations on the web. https://w3.org/blog/news/archives/6156. Date accessed: 2017-06-12.

Walker, R. and Rocha da Silva, P. (2015). Emerging trends in peer review—a survey. *Frontiers in Neuroscience*, 9:169, DOI: 10.3389/fnins.2015.00169.

Walsh, E., Rooney, M., Appleby, L., and Wilkinson, G. (2000). Open peer review: a randomised controlled trial. *The British Journal of Psychiatry*, 176(1):47–51, DOI: 10.1192/bjp.176.1.47.

Wang, W.-T. and Wei, Z.-H. (2012). Knowledge sharing in wiki communities: an empirical study. *Online Information Review*, 35(5):799–820, DOI: 10.1108/14684521111176516.

Ware, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community – results from an international study. *Information Services and Use*, 28(2):109–112, DOI: 10.3233/ISU-2008-0568.

Ware, M. (2011). Peer review: Recent experience and future directions. *New Review of Information Networking*, 16(1):23–53, DOI: 10.1080/13614576.2011.566812.

Warne, V. (2016). Rewarding reviewers–sense or sensibility? a Wiley study explained. *Learned Publishing*, 29(1):41–50, DOI: 10.1002/leap.1002.

Webb, T. J., O'Hara, B., and Freckleton, R. P. (2008). Does double-blind review benefit female authors? *Trends in Ecology & Evolution*, 23(7):351–353, DOI: 10.1016/j.tree.2008.04.001.

Weicher, M. (2008). Peer review and secrecy in the "information age". *Proceedings of the American Society for Information Science and Technology*, 45(1):1–12, DOI: 10.1002/meet.2008.14504503155.

Whaley, D. (2017). Annotation is now a web standard. *hypothes.is Blog*, https://hypothes.is/blog/annotation-is-now-a-web-standard. Date accessed: 2017-06-12.

Whittaker, R. (2008). Journal review and gender equality: a critical comment on budden et al. *Trends in Ecology & Evolution*, 23(9):478–479, DOI: 10.1016/j.tree.2008.06.003.

Wicherts, J. M. (2016). Peer review quality and transparency of the peer-review process in open access and subscription journals. *PLOS ONE*, 11(1):e0147913, DOI: 10.1371/journal.pone.0147913.

Wikipedia contributors (2017). Digital medievalist. https://en.wikipedia.org/w/index.php?title=Digital_Medievalist&oldid=779335461, https://en.wikipedia.org/w/index.php?title=Digital_Medievalist&oldid=779335461. Accessed: 2017-6-16.

Xiao, L. and Askin, N. (2012). Wikipedia for academic publishing: advantages and challenges. *Online Information Review*, 36(3):359–373, DOI: 10.1108/14684521211241396.

Xiao, L. and Askin, N. (2014). Academic opinions of Wikipedia and open access publishing. *Online Information Review*, 38(3):332–347, DOI: 10.1108/OIR-04-2013-0062.

Yarkoni, T. (2012). Designing next-generation platforms for evaluating scientific output: what scientists can learn from the social web. *Frontiers in Computational Neuroscience*, 6:72, DOI: 10.3389/fncom.2012.00072.

Yli-Huumo, J., Ko, D., Choi, S., Park, S., and Smolander, K. (2016). Where is current research on blockchain technology?—a systematic review. *PloS one*, 11(10):e0163477.

Zamiska, N. (2006). Nature cancels public reviews of scientific papers. *The Wall Street Journal*, https://wsj.com/articles/SB116663543849355795. Date accessed: 2017-06-12.