



# Progressive and degenerative journals: on the growth and appraisal of knowledge in scholarly publishing

Daniel J. Dunleavy<sup>1</sup>

Received: 31 March 2022 / Accepted: 29 September 2022 / Published online: 9 November 2022  
© Springer Nature B.V. 2022

## Abstract

Despite continued attention, finding adequate criteria for distinguishing “good” from “bad” scholarly journals remains an elusive goal. In this essay, I propose a solution informed by the work of Imre Lakatos and his *methodology of scientific research programmes* (MSRP). I begin by reviewing several notable attempts at appraising journal quality – focusing primarily on the impact factor and development of journal blacklists and whitelists. In doing so, I note their limitations and link their overarching goals to those found within the philosophy of science. I argue that Lakatos’s MSRP and specifically his classifications of “progressive” and “degenerative” research programmes can be analogized and repurposed for the evaluation of scholarly journals. I argue that this alternative framework resolves some of the limitations discussed above and offers a more considered evaluation of journal quality – one that helps account for the historical evolution of journal-level publication practices and attendant contributions to the growth (or stunting) of scholarly knowledge. By doing so, the seeming problem of journal demarcation is diminished. In the process I utilize two novel tools (the mistake index and scite index) to further illustrate and operationalize aspects of the MSRP.

**Keywords** Methodology of scientific research programmes · Sociology of science · Journalology · Scholarly publishing · Meta-science · Predatory publishing

## 1 Introduction

For the last several centuries, academic journals and their concomitant processes (e.g., editorial arbitration, peer review) have been the dominant means by which scholarship and research is appraised, validated, and disseminated (Csiszar, 2018; Porter, 1992; Smith, 2006). Indeed, many scientific discoveries are reported, critiqued, and shared

---

✉ Daniel J. Dunleavy  
dunldj@gmail.com

<sup>1</sup> Center for Translational Behavioral Science, Florida State University, 2010 Levy Ave, Building B, Suite B0266, Tallahassee, FL 32310, USA

within the pages of these publications. Einstein's "Zur Elektrodynamik Bewegter Körper" ("On the Electrodynamics of Moving Bodies"), for example, which was foundational in his subsequent development and refinement of the general theory of relativity, was published in 1905 in the influential German journal *Annalen der Physik* (Einstein, 1905). This work was later confirmed by Eddington and colleagues (Dyson et al., 1920) whose account and results of the 1919 solar eclipse were published in the *Philosophical Transactions of the Royal Society*. In the century since, publication in academic journals has dramatically increased. Today, the number of articles published per year is estimated to be around 3 million (see Johnson et al., 2018).

Though scholarly discourse has increasingly expanded outside the confines of journals (e.g., through use of preprint servers, personal and academic blogs, podcasts, etc.; see Kupferschmidt, 2020; Mollett et al., 2017; Quintana & Heathers, 2021), they have retained their dominant role within the scientific process – often being characterized as the "gatekeepers" of scientific and scholarly knowledge (e.g., Caputo, 2019; Demeter, 2020; Siler et al., 2015). However, there is substantial variability across journals (even within a single field) in terms of a journal's published output (e.g., number of volumes, issues, and articles published per year, the closed/open access status of articles), the mechanisms and processes by which articles are evaluated (e.g., open/closed publication of reviews, blinding/unblinding of reviewers and authors, data sharing policies), and the quality and content of the work published (e.g., presence and rates of selective reporting of results, data fabrication). What's more, individual journals are malleable – their internal processes and external output may change dramatically over time.

Because of this variability and the broader technological and institutional changes occurring across scholarly publishing (see generally Tennant et al., 2017), scientific and philosophical evaluation must advance beyond the traditional institutions and norms used to assess journal quality. A framework is needed to capture how and why some journals review and publish what might be called *genuine* contributions to the corpus of knowledge more frequently than others – the latter of which varyingly publish (and thus perhaps tacitly endorse) false or ambiguous claims (Camerer et al., 2018; Dunleavy, 2021; Gambrill, 2018; Open Science Collaboration, 2015). This framework will need to account for both momentary states (i.e., how a journal functions at a given timepoint) and wider trends, which extend across years and decades. Lastly, any such framework must enable direct comparison of a set of journals – since journals often compete with one another at various levels of analysis.

For instance, journals may publish peer-reviewed manuscripts that are based on suboptimal research practices (see generally Chambers, 2017; Dunleavy, 2020), fraud (e.g., Harvey, 2020; Stroebe et al., 2012), and/or plagiarism (e.g., Baždarić et al., 2012); among a host of other issues (e.g., Ioannidis, 2005).<sup>1</sup> By doing so, these

---

<sup>1</sup> Journals and their peer reviewers may also stunt growth within a field or body of knowledge by blocking (or slowing) publication of new and innovative ideas or methods, in an effort to ensure high quality control (e.g., Horrobin, 1990). O'Connor (2019) lends support to this idea, demonstrating via simulation that many scientific communities select against the conduct of high-risk/high-reward work (though she rightly notes that this varies across communities, research domains, and in relation to the expected pay-off of any given research project). Similarly, in their literature search of grant peer review, Guthrie et al. (2017) found relatively strong evidence for anti-innovation bias.

journals may hinder the development of a robust body of knowledge (Dunleavy, 2021), by introducing false, misleading, unverified, or otherwise unjustified claims into the scholarly literature (Akça & Akbulut, 2021; Akerlof & Michailat, 2019; Ioannidis, 2016; Smaldino & McElreath, 2016). The offshoot of these differences, however, is that journals can, in principle, be appraised and contrasted in accordance to how well they function in their role as gatekeepers.<sup>2</sup>

## 2 “Good” and “bad” journals

A variety of labels, designations, and metrics exist to help gauge a journal’s performance in gatekeeping or otherwise adjudicate between so-called “good” and “bad” journals.<sup>3</sup> While this is not the place for a comprehensive overview (see Tobin, 2004), a couple of brief examples will help illustrate this effort. Perhaps most notably is the commonly used journal “impact factor” (Garfield, 1972). The impact factor (IF) gives a rough approximation of how often articles from a particular journal are cited — by dividing the *total number of citations* across some period of time (e.g., 1-year, 5-years, etc.) by the *number of citable articles*<sup>4</sup> in some finite set. Conventional thinking suggests that the *higher* the impact factor (relative to a respective field) the greater the quality and prestige of a journal — the *lower* the impact factor, the lower the quality, prestige, and (perhaps) rigor (see Sternberg, 2018). Journals with the highest impact factor, in their respective field, are accordingly viewed as containing the strongest scholarly contributions. Though the IF has entrenched itself into the heart of contemporary scholarly publishing, and seemingly provides *some* utility as a metric, (Garfield, 1972, 2006; Hoeffel, 1998), its value as a measure of journal quality is questionable.

Select criticisms include, but are not limited to: Incorrectly imputing value to individual articles or authors based on journal-level metrics (i.e., journal-impact factor; JIF) (Ioannidis & Thombs, 2019), the equivocation of merit or scientific impact with IF (Dunleavy, 2022; Eyre-Walker & Stoletzki, 2013), the failure to control for self-citation practices by authors and journals (Larivière & Sugimoto, 2019), skewness in the distribution of citations among articles (Nature, 2005), the manipulation, negotiation, and general opacity of IF calculations (Brembs et al., 2013; Rossner et al., 2007), and English-language bias.<sup>5</sup> One critic (Brembs, 2018), posits that methodological rigor decreases as IF increases – a point that, if true

<sup>2</sup> This is not merely an academic exercise. As the COVID-19 pandemic has demonstrated, scholarly journals can have profound impact by informing and shaping policies that impact the health and well-being of billions of lives. Retraction Watch (n.d.) and work by Zdravkovic et al. (2020) demonstrate that even leading medical journals (e.g., *The Lancet*, *the New England Journal of Medicine*; *The Journal of the American Medical Association*) are left exposed to threats of fraud, error, and suboptimal research practices.

<sup>3</sup> As will be seen, such judgments can be made solely with non-epistemic considerations (e.g., prestige, reach) in mind.

<sup>4</sup> What is considered “citable” is often opaque and open to manipulation (as discussed by Brembs et al., 2013, p. 6).

<sup>5</sup> See Larivière and Sugimoto (2019) and Seglen (1997) for a discussion of these and other criticisms.

and generalizable, would seriously undermine use of the IF, as a proxy for quality, altogether.

In a similar vein, scholars have attempted to differentiate between quality, trustworthy, or “reputable” journals on the one hand, and untrustworthy or so-called “predatory” ones, through the creation of various whitelists and blacklists (Beall, 2010, 2014; Bisaccio, 2018; Grudniewicz et al., 2019; Laine & Winker, 2017; Teixeira da Silva et al., 2021). The underlying aim here is largely the same – journals found on blacklists are often viewed as containing weak, fraudulent, or otherwise flawed scholarship (i.e., they are “bad journals” and to be avoided). For example, blacklists (e.g., Beall’s List and Cabell’s; Bisaccio, 2018; Chen, 2019) have been used to identify journals that purportedly fail to meet professional standards<sup>6</sup> and/or are alleged to engage in exploitative or otherwise dishonest behaviors. Whitelists, such as the PubMed journal list, the Directory of Open Access Journals (DOAJ), and the Council on Publication Ethics (COPE), among others, serve as a means for communicating that a set of journals has been vetted or deemed “trustworthy” by some set of criteria or individual/group assessment (i.e., they are “good journals”). Surely these tools have helped authors detect and avoid some bad faith actors in the publishing world. However, several limitations hamper these types of lists, including:

- 1) the inability of scholars to agree upon a precise and objective definition of – or criterion for – the term “predatory” (e.g., Aromataris & Stern, 2020; Cobey et al., 2018; Grudniewicz et al., 2019; Teixeira da Silva et al., 2019),
- 2) the heterogeneity and somewhat arbitrariness of characteristics subsumed under the “predatory” label (e.g., Shamseer & Moher, 2017; Shamseer et al., 2017; Tennant et al., 2019), and
- 3) the problem of “false positive” and “false negatives” cases (i.e., mistakenly labeling a non-predatory journal as “predatory” and vice versa; Teixeira da Silva & Tsigaris, 2018; Tsigaris & Teixeira da Silva, 2021)

These issues may ultimately lead scholars to rethink the use of these labels and ensuing lists altogether (see discussion by Anderson, 2015; Kratochvíl et al., 2020; Shamseer & Moher, 2017).<sup>7</sup> Neylon (2017) has gone so far as to argue that blacklists, specifically, are “technically impossible to make work” (p. 2).<sup>8,9</sup>

<sup>6</sup> This may include lack of or inadequate peer review, insufficient expertise (or qualifications) among reviewers, inappropriately high or low article processing charges (APCs), unrealistically quick time between submission and publication, among numerous other charges.

<sup>7</sup> Similar issues beset checklist-based approaches to defining and identifying predatory journals (see Ng & Haynes, 2021).

<sup>8</sup> Other criticisms include: the pejorative or stigmatizing labeling of journals as “predatory”, discrimination of non-Western countries and countries from the Global South, the (potentially invalid) inference that poor publication practices entail that a given individual paper is necessarily also of low quality (see generally Frandsen, 2019; Kendall, 2021, p. 382), and (potential) conflicts of interest among purveyors of blacklists – particularly if they are commercial in origin (Tennant, 2020, p. 153).

<sup>9</sup> Crawford (2014) takes a somewhat different approach. He flips the burden of proof, stating that, “[j]ust because a journal exists or has a given business model or is from a given publisher doesn’t automatically make it legitimate or high quality”. (p. 1).

What is most interesting about these efforts to appraise journals (and their respective difficulties) is that they resemble and overlap with many of the core problems described by philosophers of science throughout the nineteenth and twentieth centuries – where the quest for a suitable demarcation criterion to distinguish between science and pseudoscience, the nature of scientific progress, and the sociology of science, were at the forefront of discussion and debate.<sup>10</sup> For instance, the attempt to distinguish between predatory and non-predatory (or “good” and “bad”) journals (Siler, 2020) resembles Popper’s (1959/1968) classical formulation of the demarcation problem (i.e., how to distinguish science from pseudoscience). Contemporary questions about the role of journals and publishers in the development and growth of scholastic knowledge (see Dunleavy, 2021; Lock, 1985, especially Chapter 6) mirror investigations into the concept of, and theories about, scientific progress (e.g., Lakatos & Musgrave, 1970; Laudan, 1977). Finally, the self-governance of scholarly communities (including the reform of journal publication practices and standards; e.g., Dunleavy & Hendricks, 2020; Hardwicke et al., 2020) have links to broader inquiries into the workings, policing, and self-correction of scientific communities (e.g., Kuhn, 1962; Merton, 1973; Zuckerman & Merton, 1971). Because of these parallels, it may be fruitful to draw upon the philosophy of science to illuminate aspects of these current debates in journalology (i.e., the study of publication practices) and specific attempts to appraise journal quality.

In this article, I present a novel lens by which we can evaluate scholarly journals – one that helps capture its dynamic nature (i.e., that journal policies, functions, and published output are not fixed, but diverse, malleable, and ever-changing).<sup>11</sup> Specifically, I draw on the work of Imre Lakatos and his *methodology of scientific research programmes* (MSRP; Lakatos, 1968b, 1970, 1978b). I argue that his classifications of “progressive” and “degenerative” research programmes – and their general features – can be analogized and repurposed for the evaluation of scholarly journals. In doing so, I argue that this alternative framework resolves *some* of the flaws of current approaches discussed above and offers a more considered evaluation of journal quality – one that helps account for the historical evolution of journal-level publication practices and consequent contributions to the growth (or stunting) of scholarly knowledge.<sup>12</sup>

I begin by introducing Lakatos’s MSRP and its associated terminology. Next, I discuss how it can be repurposed for the task of evaluating scholarly journals. I note here that Lakatos’ philosophy need not be the “final word”, or even a correct account (see Larvor, 2006, especially p. 715; and Cohen et al., 1976) of the philosophy and

<sup>10</sup> The edited volume by Lakatos and Musgrave (1970) provides a good overview of work in this period. See also work by Campbell (1987), Kuhn (1962, 1970), Lakatos (2012), Popper (1963, 1970), Reichenbach (1938), and Scheffler (1963).

<sup>11</sup> These changes can be brought about internally or in response to external pressures (e.g., McNutt, 2016).

<sup>12</sup> To be clear, I do not view this effort as replacing the use of the IF or blacklists/whitelists and other contemporary approaches. Rather, I view it as broadening the scope and focus of journal appraisers. Tools like the IF, despite their flaws, may yet be but one useful form of information to be integrated into these appraisals. This point is briefly explored in more detail below.

sociology of science, to be a fruitful tool in this respect.<sup>13</sup> After all, even flawed philosophies can still be useful.

Having introduced the MSRP, I then attempt to operationalize its components – relying on two metrics: the “mistake index” (MI; see Margalida & Colomer, 2015) and the “scite index” (SI; see scite, 2019b) – to help illustrate empirical and theoretical features of the progressive and degenerative classifications. These metrics provide preliminary criteria for assessing whether a journal’s articles – as a function of its internal editorial policies and practices (e.g., peer review) have caused it to enter a progressive, degenerative, or stagnant phase. Finally, I discuss how such an approach could complement or supplant contemporary (if flawed) evaluative schemes, such as the impact factor and the development of predatory lists, described above, and outline an agenda for future scholarship on this topic.

### 3 Lakatos and the methodology of scientific research programmes

Imre Lakatos (1922–1974) was a Hungarian-born philosopher of science. Shortly after World War II, he immigrated to England, where he completed his PhD in philosophy (1961 – King’s College Cambridge). In 1960, he started teaching at the London School of Economics. As a philosopher Lakatos had a number of interests, but primarily centered his focus on the philosophies of mathematics and science (see Lakatos, 1976, 1978a, 1978b). Lakatos would go on to publish on topics such as the problem of demarcation, theory-change, and induction, as well as the growth of scientific knowledge, until his sudden death in 1974.<sup>14</sup>

Within the philosophy of science, he is perhaps most noted for his development and articulation of the methodology of scientific research programmes – a framework which he viewed as an advance from the work of his contemporaries Thomas Kuhn and Karl Popper (Lakatos, 2012, p. 23).<sup>15</sup> Kuhn (1962, 1970), to his end, attempted to capture the process of scientific change and the workings of scientific communities with his concept of the “paradigm” and the historically-driven descriptions of “normal” and “revolutionary” science. Popper (1959/1968, 1963), on the other hand, attempted to justify a rational, deductive approach to scientific discovery, via his method of “conjectures and refutations” – one that embraced a Humean skepticism towards induction and challenged the perceived difficulties inherent in the contemporary tradition of the logical positivists. Though Lakatos was critical of both of these attempts, he was sympathetic to each as well. Consequently, features of both Kuhn and Popper can be found in his work and general philosophical outlook.<sup>16</sup>

<sup>13</sup> This point is also noted by McCloskey with specific regard to the MSRP (1993, p. 230).

<sup>14</sup> See Larvor (1998) and Musgrave and Pigden (2016) for a deeper review of his personal life and work.

<sup>15</sup> Whether, in fact, his MSRP was an advance is certainly up for debate. See Bartley (1976), Blaug (1976), Kuhn (1980), de Marchi and Blaug (1991), Popper (1982), and Worrall (1978) for in-depth, and sometimes heated, critique of this claim.

<sup>16</sup> Similarly, Kadvaný (2001) characterizes the development of the MSRP as being a “synthesis and culmination of a sequence of successively more sophisticated and powerful methodologies” (p. 157), including, but not limited to those of Popper and Pierre Duhem – and perhaps implicitly with figures such as Marx and Hegel (p. 2; see also Motterlini, 2002).

The MSRP can be viewed as an attempt to bring aspects of the rationality (or “logic”), practice, and history of science into a comprehensive framework. As will be described below, it touched not only on matters of theory-testing and evidential support (e.g., Popperian falsification and corroboration), but also the methodological and pragmatic decisions made by scientists (and scientific communities) about how and when to pursue or withdraw from lines of inquiry (e.g., Kuhn’s “normal” and “revolutionary” science and puzzle-solutions) – as well as the historical track record (successes and failures) of theories and their conceptual rivals.

Under Lakatos’ MSRP (1968b, 1970), a research programme has four key components, the: 1) unchanging or irrefutable “hard core”, 2) a “protective belt”, 3) the “negative heuristic”, and 4) the “positive heuristic”. Ideally, research programmes will all have these four components, though in practice, some components have been more clearly articulated than others. The hard core comprises the essential elements of a programme. It is what we might call the “lead idea” (Larvor, 2006) or “central principles” (Hacking, 1983). In Newton’s theory of gravitation, the hard core contains his three laws of motion and his one law of gravitation. Together, these four laws serve as the foundation for which empirical observations are predicted and explained. Accordingly, they are not to be questioned or modified.<sup>17</sup> The protective belt, in contrast, brings a sort of stability to the programme. It consists of the set of *auxiliary* hypotheses that further articulate and support the underlying theory (i.e., the hard core), and most importantly, shield it from potential falsifiers or other empirical anomalies. In other words, these hypotheses serve an instrumental purpose (i.e., to “protect” the hard core). They are disposable, eventually becoming modified or replaced by new auxiliary hypotheses. Again, using Newton’s programme as our example, the protective belt would consist of his theories of atmospheric refraction and geometrical optics (Lakatos & Zahar, 1975; Lakatos, 1971b) – theories which can be tweaked or supplanted in the face of inconsistencies between predictions made from Newton’s four laws and subsequent observations. Finally, are the negative and positive heuristics. Heuristics are the implicit and explicit methodological rules and beliefs which constrict and guide *the behavior* of scientists working within the programme. The negative heuristic tells the scientist what *not* to do (e.g., “...not to tinker with the hard core...”, Chalmers, 1999, p. 133), while the positive heuristic tells the scientist where to focus their attention – which provides what Hacking (1983) describes as a “...ranking of [scientific] problems” (p. 117) to work on. By doing so, the positive heuristic prevents the scientist from being distracted by empirical anomalies or otherwise unimportant or distracting issues. Together, the two heuristics help facilitate research within the programme.

Having sketched out some basic features of a Lakatosian research programme, we can now turn to how they are classified and appraised. Lakatos calls programmes which are performing well “progressive” and those that are failing or otherwise languishing as “degenerative”.<sup>18</sup> A programme is considered progressive when it makes

<sup>17</sup> One might consider them to be (in some sense) axiomatic.

<sup>18</sup> These are categorical terms with dimensional properties – that is, something can be more, or less, progressive/degenerative. Lakatos (1971a, footnote 36, p. 125) uses the notion of *ad hocness* to describe and categorize the various ways in which a programme may degenerate (see Zahar, 1973a for an exposition and refinement of the classifications of: *ad hoc*<sub>1</sub>, *ad hoc*<sub>2</sub>, and *ad hoc*<sub>3</sub>).

novel predictions and has (at least some of) these predictions confirmed/corroborated. These predictions should be consistent with the programme's positive heuristic. The more a programme achieves these tasks, compared to its rivals, the more progressive it is deemed.<sup>19</sup> In contrast, a degenerative programme fails to make novel, confirmed predictions – or does so in spite of its own flawed internal logic. Rather, a degenerating programme's development (e.g., modification of auxiliary hypotheses) occurs not due to its own internal success, but “in response to external criticism” (Larvor, 2006, p. 713) or in reaction to the success(es) of its competitors. For Lakatos, Newton and Einstein's theories were paradigmatic cases of progressive programmes, with socioeconomic Marxism and Freudian psychoanalysis epitomizing degenerative ones (Lakatos, 2012). More contemporary examples of progressive programmes might include investigations into the effects of childhood adversity and social deprivation on the development of later psychopathology (e.g., McLaughlin & Hatzenbuehler, 2009; Varese et al., 2012; Wickham et al., 2014), the HIV-AIDS hypothesis (Barré-Sinoussi et al., 1983; Gallo, 1991; Gallo & Montagnier, 2003), and more recently the CRISPR gene editing programme within molecular biology (Hsu et al., 2014) on the one hand, and research investigating genetic or monoamine-deficit causes of depression (e.g., Border et al., 2019; Cai et al., 2020; Healy, 2015; Lacasse & Leo, 2005), the drug-induced or “chemical” hypothesis of AIDS (see generally Ellison & Duesberg, 1994; Duesberg, 1992, 1996a, 1996b; Duesberg et al., 2003; Ellison et al., 1995), and perhaps, more broadly, collections of programmes – such as those found in some areas of nutritional science nutritional science (e.g., Ioannidis, 2013; Taubes, 2007) on the other.

A final point on Lakatos' MSRP is worth stressing here. In Lakatos' philosophy, progressive and degenerative appraisals are contextual, contrastive, and impermanent – features which are distinct yet intertwined. A programme is progressive or degenerative with respect to how it fares *relative to its rivals* (e.g., does *Programme A* out predict/perform *Programme B*?) at a given point in time (or across some finite timeframe). Moreover, these designations are not fixed – that is, a progressive programme can begin to fail in time, while a degenerating one may still yet recover and flourish. In Lakatos's (1971a) own words:

“One must realise that one's opponent [rival programme], even if lagging badly behind, may still stage a comeback. No advantage for one side can ever be regarded as absolutely conclusive. There is never anything inevitable about the triumph of a programme. Also, there is never anything inevitable about its defeat...The scores of the rival sides, however, must be recorded and publicly displayed at all times.” (p. 101)

Accordingly, the MSRP's strengths reveal its purported limitations. That is, it is often argued that Lakatos' MSRP permits a contextual and historically-driven assessment of scientific research (appraisal), but does not allow for definitive conclusions or recommendations (e.g., what lines of research to pursue or abandon)

<sup>19</sup> One might also consider traits other than predictive value in the assessment of a programme—such as its explanatory value or increasing precision in measurement—traits which support, but may be independent of the ability to predict.



to be made.<sup>20</sup> We will return to these issues in more detail later. For now, we will attempt to motivate the case for applying the MSRP to scholarly journals.

#### 4 Progressive and degenerative journals

Though Lakatos's MSRP (1968b, 1970) largely dealt with the empirical sciences like physics, chemistry, and astronomy – fields with highly mature theoretical content (e.g., Newton's theory of gravitation; Bohr's model of the atom, Copernicus' heliocentric model of the universe; Lakatos & Zahar, 1975; Musgrave, 1976b; Zahar, 1973a, 1973b) – there is no clear rationale for precluding its use in describing other areas of inquiry. Lakatos himself (1971a) makes this point clear, when he states that, “[T]he methodology of research programmes may be applied not only to norm-impregnated historical knowledge but to *any normative* knowledge, including even ethics and aesthetics.” (p. 132, emphasis added). This point is affirmed by Kadavy (2001), who argues that, “Lakatos intends the methodology of scientific research programmes as a general theory of criticism...” (p. 216). Such non-empirical, quasi-empirical, and extra-scientific applications of the MSRP are exemplified by Lakatos himself, in his framing of Rudolf Carnap's system of inductive logic as a degenerating research programme (Lakatos, 1968a; see also Groves, 2016), his reappraisal of the methodology of “proofs and refutations” as a progressive programme within mathematics (Lakatos, 1970, p. 180, footnote 2), and in his discussion of Popper's philosophical account of the growth of scientific knowledge (Lakatos, 1970, p. 180, footnote 2; see also Lakatos, 1978b, Chapter 3 and Feyerabend, 1974). His contemporaries demonstrate the utility and flexibility of the MSRP as well. For example, D'Amour (1976) in his discussion of *act utilitarianism* within the field of ethics (see particularly pp. 89–93; and Alfano, 2013), Koertge in her discussion of the philosophical school known as logical empiricism (1972), and by Popper (1974) in his (initial) characterization of evolutionary theory and Darwinism as *metaphysical* research programmes (pp. 118–121 and 133–143).<sup>21</sup>

Even Feyerabend, Lakatos's close friend and intellectual antagonist (Motterlini, 1999), floats the possibility of *theological*<sup>22</sup> research programmes (Feyerabend, 1975, p. 15), if we are by necessity to consider programmes within the context of their contemporary rivals, scientific and otherwise. These examples, along with later applications of the MSRP in fields far removed from philosophy and the hard sciences, such as economics (Latsis, 1976), developmental psychology (Phillips, 1987; Phillips & Nicolayev, 1978), intelligence research (Urbach, 1974a, 1974b),

<sup>20</sup> The distinction between *acceptance* and *pursuit* is discussed in detail by Barseghyan and Shaw (2017). The authors posit that Lakatos was primarily concerned with theory acceptance, while Feyerabend was mainly concerned with pursuit. In practice, Lakatos seemed to hold a (implicit) blend of both stances (see pp. 8–9).

<sup>21</sup> See Elgin and Sober (2017) for critical discussion of Popper's views.

<sup>22</sup> Note, here Feyerabend was discussing scientific practice in the sixteenth and seventeenth centuries. For a more modern discussion in this vein, see discussion by Murphy (1990, especially Chapter 4), Hefner (1993), and Lorrimar (2017).

and international relations (Elman & Elman, 2003), demonstrate the flexibility of the MSRP and give precedent for its application (and potential utility) within the broader areas of journalology, meta-research, and meta-science discussed here.<sup>23</sup> Since MSRP attempts to capture the rationality of scientific practice and the growth of knowledge, it could conceivably be employed in the context of scholarly journals, where the practice of scholarly publication – *vis-à-vis* journals, provides a (ostensibly) rational means for assessing new ideas and evidential claims and validating and disseminating new knowledge.

#### 4.1 Journals as research programmes

Journals are composed of many different parts and processes. There is the obvious published output, namely, journal articles – which may, depending on the journal, include empirical studies, conceptual pieces, editorials, commentaries and opinions pieces, letters to the editors, book reviews, and so on. Journals are also comprised of stakeholders, including but not limited to the editor(s)-in-chief, editorial board, reviewers, the authors who submit to and publish in the journal, and (perhaps also to some extent) the journal's broader readership. Still further, are the standards and policies by which the journal performs its activities. This may include its stated "aims and scope", instructions to authors and reviewers (i.e., the basic standards submitted manuscripts must conform to and standards by which reviewers are to adhere in evaluating those manuscripts), the internal/external processes by which manuscripts are solicited, evaluated, published and rejected (e.g., the number of referees utilized, anonymization of authors/reviewers, criteria for determining acceptability, etc.), and the output of those processes (e.g., reviewers reports), among other things. These components are brought together within the journal's parameters and are materialized.<sup>24</sup>

Some of the features of journals can be readily redeployed within the four parts of the Lakatosian framework. Recall, these are the 1) hard core, 2) protective belt, 3) negative heuristic, and 4) positive heuristic. The negative and positive heuristics are readily identifiable. In scientific research programmes these are the rules and beliefs that guide the behavior of scientists and drive experimentation. The analogous features in this context would be the tacit and explicit rules, norms, and policies that guide the journal's stakeholders (namely, but not exclusively, the editorial board and peer reviewers). Examples abound. The *Public Library of Science* (PLOS) family of journals, for instance, requires "authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication" (PLOS One, 2019), barring certain ethical and legal exceptions. Another example comes from the field of public health. Kenneth Rothman (1998), former Editor-in-Chief of the prominent journal *Epidemiology*, notably discouraged authors from using *p*-values. While acknowledging that they

<sup>23</sup> At the time of this writing, Fernandez-Cano (2021) made a passing reference to Lakatos's methodology as potentially applying to the area of scientific publishing (p. 3677).

<sup>24</sup> Influential journals like *Science*, *Nature*, *JAMA*, *BMJ*, and *Cell* typify the description above.

are sometimes productively used, he noted that, "...we prefer that  $p$ -values be omitted altogether, provided that point and interval estimates, or some equivalent, are available." (p. 334; see Trafimow, 2014).<sup>25</sup> More generally, each journal's "Aims and Scope" description helps guide editors and reviewers in determining the "fit" between a manuscript's content and the journal's overarching mission or purpose. Together, these heuristics and rules guide a journal's stakeholders towards manuscripts that it finds of value; and away from ones it disvalues,<sup>26</sup> while constraining or nudging authors to submit content which coheres with the form, standards, and objectives of the journal.<sup>27</sup> Note here, that this kind of valuing might ideally be said to refer to manuscripts that are additive to our knowledge base, but this is not necessarily so (see generally Else, 1978, pp. 269–270). Journals can, after all, value "flashy" papers and results or other instrumental and non-epistemic characteristics over, or in addition to, seeking scholarship which promotes substantive research and the search for truth (see generally Brembs, 2018; Nosek et al., 2012; Serra-Garcia & Gneezy, 2021; Smith, 2006).

Having identified how heuristics come into play within journals, we can now focus on the other two components of the MSRP, the hard core and protective belt. This effort is admittedly less clear, but we can sketch out some possibilities. In traditional scientific research programmes, the hard core consists of the central principles or ideas. The protective belt consists of a (relatively) disposable set of auxiliary hypotheses. The latter support the former and serve to ward off empirical anomalies and other threats. What then, is the central, fixed, "core" of a journal? Certainly, no single or finite (yet everchanging) number of articles could be said to make up the core of a journal. Articles are malleable. They can be corrected, retracted, and critiqued. Nor is the core represented by the stakeholders and staff of a journal, for they too change with the passing of time. I argue that the hard core could, perhaps, be best represented by something slightly more abstract – what might be understood as the overarching aim (or "ethos") of the journal.<sup>28</sup>

<sup>25</sup> In a similar vein, see JAMA (2021) on the use of "causal language" in interpreting study findings.

<sup>26</sup> In the case of *Epidemiology*, this might have meant that editors and peer reviewers explicitly rejected or strongly critiqued manuscripts using  $p$ -values.

<sup>27</sup> It is also the case that journals influence subsequent research by virtue of what they choose to publish. Individual papers, which may have the tacit endorsement of the journal in which they are published, may promote, inhibit, or direct future research efforts. This occurs in at least two ways – 1) published papers serve as precedent for what is acceptable to publish within a specific journal and 2) papers may contain recommendations for what research and scholarship should be prioritized or admonished going forward. For example, a highly cited and disseminated study by Border and colleagues (2019), published in *The American Journal of Psychiatry* (one of the leading psychiatric journals in the world and the official journal of the *American Psychiatric Association*) forcefully challenged the validity of an entire program of research in behavioral genetics. The authors close their piece by stating, "[W]e conclude that it is time for depression research to abandon historical candidate gene and candidate gene-by-environment interaction hypotheses." (p. 386). [The author would like to thank Reviewer #2 for highlighting the general point made here].

<sup>28</sup> A similar point is made by Lakatos' former student Donald Gillies (2002): "Returning to Lakatos' theory of scientific research programmes, my view is that it forms part of the theory of the growth of science and mathematics. With a single modification, I accept this theory as a valid and important part of this branch of the philosophy of science and mathematics. ... I prefer to substitute for the concept of the *hard core* of a research programme, that of the *aim* of a programme." (p. 19; emphasis in original).

A few real-world examples will help illustrate what I mean by aim. The *New England Journal of Medicine* (henceforth *NEJM*; n.d.), states that “Our mission is to publish the best research and information at the intersection of biomedical science and clinical practice and to present this information in understandable, clinically useful formats that inform health care practice and improve patient outcomes”. *Psychological Science* (henceforth, *PS*; n.d.), one of the foremost psychology journals, describes itself as publishing, “cutting-edge research articles and short reports, spanning the entire spectrum of the science of psychology. This journal is the source for the latest findings in cognitive, social, developmental, and health psychology, as well as behavioral neuroscience and biopsychology”. *Social Work Research* (henceforth *SWR*; n.d.) purports to publish, “...exemplary research to advance the development of knowledge and inform social work practice”. Each journal is ostensibly a venue in which to publish rigorous and/or novel research, which helps to advance its discipline’s knowledge base and/or improves clinical practice. Regardless of whether these journals are successful in these efforts, we can see that something lies at the center of each journal – something that would fundamentally change the journals themselves if altered – regardless of its physical content (i.e., stakeholders, published output).<sup>29</sup> Put differently, at its core, a journal consists of an overarching scholarly ethos (i.e., the journal’s primary aim), whose internal sorting processes (editorial decision-making and peer review) are guided by a set of positive and negative heuristics. This results in submitted manuscripts being validated and assimilated into the journal (i.e., “published”) or rejected and discarded. When this underlying ethos changes (e.g., in response to external pressures or internal dysfunctions), so too has something fundamental at the journal’s core.

A journal’s published articles (as objects) could be said to comprise the protective belt. Articles published in the *NEJM*, for example, support (in varying ways) the journal’s hard core.<sup>30</sup> Articles that successfully sustain this core help foster a progressive programme. Inasmuch as a journal’s core is concerned with scientific/epistemic progress, this may, among other things, mean that the hypotheses posited within empirical articles are subject to severe tests (e.g., Mayo, 2018), which are subsequently confirmed by the results of the performed analyses, and are later corroborated and/or replicated in future studies. Ones that do not,<sup>31</sup> contribute to the degeneration of a

<sup>29</sup> Take for example what *PS* would be if it shifted its focus toward publishing only articles related to Freudian psychoanalysis. Not only would we say that the content of the journal (i.e., its articles) has changed, but something fundamental about the journal itself. It no longer maintains its core principles, tenants, aims. It has lost its central ethos.

<sup>30</sup> Support can be cashed out in many ways (e.g., predictive power, explanatory power, corroboration of theory, etc.) – which are explored in the next section. For now, it will suffice to recognize that the articles, as objects, exist separate from the journal’s core.

<sup>31</sup> This may be a due to the choice of hypotheses (and attendant background theory), the unreplicability of findings, lack of methodological rigor, or among other things, to the use of questionable research practices (QRPs; which include, but are not limited to, “p-hacking” [e.g., collecting data until results are statistically significant], “HARKing” or hypothesizing after results are known, and selective reporting or “cherry picking” of statistically significant results. See discussions by Chambers, 2017; Ioannidis, 2016; and Simmons et al., 2011). This is further compounded by vague or faulty heuristics that lead journal editors and reviewers to prefer unreplicable or “sloppy” studies over those more rigorously designed and executed (see generally Goldacre, 2008; Smaldino & McElreath, 2016).

programme. Degeneration, then, can be characterized<sup>32</sup> by increasing rates of corrections, retractions,<sup>33</sup> and outside critique (e.g., refutation). While these criticisms are very often directed at the article(s) and its authors,<sup>34</sup> critical attention may inevitably shift towards the journal itself. As these criticisms add up, the journal may fall into disrepute (though this is not strictly true; see Fanelli, 2013).

Having described the MSRP and explicated its components in the context of scholarly journals, we can now attempt to describe its potential application. To do so, I will draw on two formalized metrics, the “mistake index” and the “scite index”. I have decided to emphasize these two metrics for illustrative purposes and because I feel they help capture (in their own limited ways) empirical and theoretical contributions of articles and more broadly the journals in which they are published. These metrics provide preliminary criteria for assessing whether a journal’s articles have caused it to enter a progressive, degenerative, or stagnant phase. However, in practice, the appraisal of research programmes will very likely rely on a number of metrics and assessments (epistemic and non-epistemic; empirical and metaphysical).<sup>35</sup>

## 4.2 The mistake index (MI)

The mistake index (MI) is a relatively novel tool, put forth by Margalida and Colomer (2015) as a measure for reviewer/editor effectiveness. In its simplest form, the MI is calculated by taking the annual number of items published in a journal (or by a publisher) and dividing by the total number of corrections (i.e., erratum/corrigendum). This provides a standardized index that can be compared from year-to-year and/or across journals. The higher the MI, the better a journal’s performance (generally speaking). To use a toy example, a journal that published 100 items in a given year, alongside ten corrections [i.e., 100/10], would have an MI of 10. Of course, as with other tools such as the IF, a journal’s MI can be calculated for any finite period of time (e.g., 5-year period, lifetime), and need not solely be done for a single year. Margalida and Colomer (2016) have since extended and refined this proposal, delineating between indices calculated according to total items published (described above), which they call the Mistake Index Total (MIT) and that calculated on the basis of total *papers* (i.e., articles)

<sup>32</sup> Of course, we may choose other empirical and non-empirical standards for judging progression/degeneration (e.g., functionality of a journal’s peer review process). Plausibly, one might even explicate progressiveness/degeneration in somewhat unsavory ways (e.g., the extent to which an exploitative journal exploits authors and institutions).

<sup>33</sup> Retractions, it must be noted, arise for many different reasons (see Andersen & Wray, 2019; Wray & Andersen, 2018). Here, I am primarily focused with retractions that are made on the basis of fraud, plagiarism, egregious error in data analysis, etc. (see these guidelines by the *Committee on Publication Ethics [COPE]* for a helpful overview of when a manuscript should be retracted). Nevertheless, as retractions accrue, even if due to non-epistemic reasons (e.g., sociopolitical pressures), and especially if initiated by external forces (rather than originating from a journal’s internal self-correcting processes), it will typically be reflected by an increasingly negative appraisal a journal’s functioning.

<sup>34</sup> See the critique by Wagenmakers et al (2011) of Bem’s (2011) famous study on “psi” phenomena or the infamous piece by Wakefield et al. (1998), critiqued by Godlee (2011).

<sup>35</sup> What these are will undoubtedly need to be decided by the broader scientific and scholarly community.

published – the Mistake Index Paper (MIP). In the following examples, all references to the MI will refer to this latter calculation.

The MI is used here for two purposes. First, it serves as a proxy of reviewer/editor effectiveness.<sup>36</sup> I say proxy because, under ideal circumstances, this would simply reflect how well a journal catches mistakes, errors, omissions, etc. with higher rates indicating increasingly suboptimal functioning of review processes. However, since journals presently lack incentive to correct published mistakes, more rigorous journals may have higher rates of mistakes relative to their (less self-critical) peers. Nevertheless, the MI still serves as a surrogate for quality. This is because a high MI can be indicative of nothing but low quality and performance.<sup>37</sup> Second, the MI captures something about a journal's empirical and theoretical contributions. A journal whose articles often contain errors (especially severe ones) are not supporting or protecting its hard core (e.g., to publish the best research on 'x'), nor reliably generating confirming evidence. Consequently, a journal who has a high MI, relative to its peers (particularly when examining a set of journals with rigorous self-assessment and post-publication peer review) – or an increasing MI, relative to itself, over time – can be characterized as being stagnant or (in extreme cases) degenerative.

### 4.3 The scite index (SI)

*scite* is a deep learning-based platform developed by researchers and academics from across the sciences. *scite*, among other things, extracts citations from published articles and categorizes them based on how they are discussed within the text of the paper (i.e., *supporting*, *contrasting* [i.e., disputing], or *mentioning* – the latter being a neutral descriptor). This allows readers (and authors) insight into how supported a particular article or claim is, within the scholarly corpus. This approach has since been quantified. The scite index (SI) provides a standardized measure by which one can evaluate the quality of scholarly publishing (e.g., at the journal-level). To calculate the SI for a given journal, one would take the number of supporting citations, across a given time period, and divide by the number of supporting and contrasting citations. For example, as of the year 2019, a given journal's articles (say, journal "X"), has had 95 supporting citations and only 5 contrasting citations. This would mean journal X has a SI of  $0.95$  [ $95/95 + 5$ ]. As

<sup>36</sup> Mistakes can be further classified in terms of their severity: mild, moderate, severe. This final category includes mistakes that result in "important modifications in figures, results or conclusions, [or] implying major changes (i.e., figure substantial modifications, re-interpretation of the results, clarification of the conclusions)" (see Margalida & Colomer, 2016, p. 4). Notably, in its current form, the MI (as applied by its authors) does not include mistakes or corrections associated with retracted works.

<sup>37</sup> Consider an analogy: Reporting of violent crime in a community may be an indirect proxy of total violent crime. Differences between actual crime and reported crime may occur due to hesitancy to report, inadequate policing, etc. As such, a community with low report rates may be a peaceful place to live or not – it may still in fact have high crime rates, relative to its peers (they are just unreported). However, a community with very high rates, due to adequate reporting and policing is nevertheless a potentially dangerous place to live, even if those reports occur, in part, due to hypervigilance of community stakeholders. Similarly, a low MI does not entail that a journal is without mistakes. But a journal with a high MI cannot be said to be a reliable and well-functioning one.

with the MIP above, a SI can be calculated for a given journal for any time interval (e.g., a 1-year period, 5-year period, lifetime). This permits both internal evaluation of a journal’s quality across time and for its comparison with peers (Nicholson, 2021; Nicholson et al., 2021; scite, 2019a).

The SI is used here for two purposes. First, it serves as a proxy for the corroboration of findings (i.e., results) and claims (e.g., hypotheses/theory) made within articles. Again, I say proxy because supportive citations may reflect well-confirmed findings, but this is not necessarily so. A paper, author, or journal with a high SI score may merely reflect groupthink among a network or field of scholars or a strong, yet misplaced (or premature), scientific consensus. Nevertheless, the SI serves as an important indicator of journal quality. Findings that are not (or never) supported in subsequent literature will (likely) not be accepted within the broader scholarly community. Second, the SI captures something about a journal’s empirical and theoretical contributions – that is, ostensibly, well-confirmed and robust findings will (inevitably) have higher SI scores than those which do not – whereas those that are disconfirmed, are inconclusive, or controversial will seemingly have middling or low scores. Consequently, a journal who has a high SI, relative to its peers – or an increasing SI, relative to itself, over time – may be characterized as being progressive. Lower or decreasing SI scores may reflect a stagnating or degenerating journal.

Ideally, the MI and SI could be used in together, in a complementary fashion – or paired with other proxies for functioning, such as comments posted after publication (i.e., post-publication peer review; see Bordignon, 2020). This would help gauge the reliability of published output as well as its epistemic and social value of specific journals to the scientific community.

Table 1 (below) is based on the metrics described above and provides preliminary standards for appraising journal functioning. A journal that has a low rates of severe MI and high SI can tentatively be labeled as *progressive*. Such a journal, under ideal conditions, is validating and disseminating research and scholarship which is both reliable and highly confirmed. A journal which has a high MI (particularly moderate and severe ones), but still produces work that corresponds with a high SI should be viewed more skeptically (*Stagnating*). Published output from this type of journal may be less reliable (though again, here, context is key – as a high MI could reflect greater editorial attention – for instance, when rates of MI reflect mild or moderate errors). A journal which has low SI and high MI may be classified as degenerating (*degenerative<sub>1</sub>*). Such a journal is producing work that is unconfirmed or disconfirmed and is of questionable reliability. Even if the high MI reflects greater editorial functioning (e.g., by correcting or preventing severe errors), the journal is of questionable utility if its contributions to the scholarly corpus lack epistemic value. Lastly, we have journals which have low MI and low SI (*degenerative<sub>2</sub>*). A journal

**Table 1** Preliminary taxonomy for appraising scholarly journals

		<i>Mistake index (MI)</i>	
		<i>High MI</i>	<i>Low MI</i>
<i>Scite index (SI)</i>	<i>High SI</i>	Stagnating	Progressive
	<i>Low SI</i>	Degenerative <sub>1</sub>	Degenerative <sub>2</sub>

with these features is ostensibly reliable, but produces unconfirmed or disconfirmed scholarship – leading one to ask what value the journal has. In absence of any evidence for self-correction, it warrants questioning whether its editorial standards and practices are unable or unwilling to solicit, appraise, and produce work which generates novel and confirmable findings.

## 5 Caveats and limitations

There are several limitations with the proposal discussed here which must be noted. Many of these problems haunt Lakatos' original formulation (see Feyerabend, 1970, 1980). Foremost, the MSRP does not offer hard and fast rules for pursuit at the level of the individual. Though the progressive and degenerative designations have utility in appraisal and assessment of theories and journals, they cannot provide definitive guidance (to the individual researcher, scholar, scientist) about which to support, endorse, avoid, or dismiss in the future. Rather, the MSRP – applied within the context of scholarly journals – is useful for overall appraisal and this assessment can then be used to inform pursuit within scholarly communities. Put differently, for the individual, it is primarily a *descriptive* and “backward-looking” (Hacking, 1983) framework – offering a pragmatic set of tools for future consideration, rather than a purely *prescriptive* one. After all, there are valid reasons that an individual may provide for backing (i.e., publishing within, reviewing for, contributing to) a seemingly degenerating journal or for avoiding a progressive one (Feyerabend, 1970, 1980).<sup>38</sup> Reflecting on replies from his colleagues, Lakatos (1971b) notes:

"The arguments my critics produce have made me realise that I fail to stress sufficiently forcefully one crucial message of my paper [1971a]. This message is that my ... ‘methodological rules’ explain the *rationale* of the acceptance of Einstein’s theory over Newton’s, but *they neither command nor advise* the scientist to work in the Einsteinian and not in the Newtonian research programme." (p. 174, emphasis added)

Despite these faults, degenerating programmes have instrumental value by applying direct or indirect pressure(s) to competing research programmes and by offering guidance for scholarly communities (as well as government agencies, funders, etc.). That latter can readily employ these designations as a rational guide about what to justifiably endorse or withhold resources from (Lakatos, 1970, p. 157, footnote 1; see also elaboration by Musgrave, 1976a, 1978).

<sup>38</sup> In some ways, this limitation of the MSRP (at the individual level) can be reframed as a strength (at the community level). Feyerabend’s (1970) *principles of tenacity* (roughly, that we needn’t quickly reject or abandon a programme that is struggling) and *proliferation* (in which we may permit the growth of theories/programmes [or journals that employ a diverse array of standards and practices]) enable us to rationally stick to theories (or journals) – exactly because they can improve over time and resolve long-standing difficulties (for discussion see Farrell, 2003, especially Chapter 8). In the context of scholarly publishing it may make sense for there to be an assortment of journals, in a given field or area of study even if some are stagnating or degenerating – of which the scholarly community can rationally invest or divest resources and effort to (e.g., by submitting manuscripts, volunteering to peer review, etc.).



The inability of the MSRP to providing absolute certainty (i.e., that past performance is a fallible indicator of future success) is readily addressed. As Motterlini (1995) notes:

“[T]o the question of whether we are able to supply any practical indication on the basis of appraisals which are concerned with past performances, we can now give a positive answer. ... Naturally, it is not the lot of humans to know the future – and the question whether past performance is a reliable guide to the future may be addressed to all methods of appraisal –, but this does not mean that any research programme is as “promising” as another”. (pp. 6-7).

In other words, the fallibility of our inferential tools (e.g.,  $p$ -values, Bayes factors in statistical inference), at the individual level, do not preclude their valid uses at the broader level of the scientific community.

Other limitations include the difficulty in fully operationalizing the “progressiveness” and “degeneration” labels (concepts that surely exist on a spectrum), the actualization and implementation of the MSRP into journalology and meta-research (i.e., whether the MSRP can be fully refashioned<sup>39</sup> for this purpose), and the ever-growing set of alternative methods of journal appraisal (e.g., Teixeira da Silva et al., 2021).<sup>40</sup> These have been highlighted here and will be addressed in future discussions of the proposal.

What is gained by this exercise is an ability to capture journal change over time (see generally Lock, 1989) – something that is largely missing<sup>41</sup> from current, static designations, such as whitelists and blacklists<sup>42</sup> and that is understood too narrowly when applied solely within individual metrics (e.g., JIF). After all, what is a “high impact”, “good”, or “legitimate” journal today, may not be tomorrow – and one that is “low impact”, “bad”, or “predatory” can ultimately change the way it performs, re-establishing (or achieving for the first time) reputability and legitimacy. This is particularly true given the shock of the COVID-19 pandemic to the publication landscape (Dunleavy et al., 2020) which has left consumers of scholarship drowning in a sea of (mis)information (Correia & Segundo, 2020; Dunleavy & Hendricks, 2020) and brings questions about the quality of published work in even leading journals (Alexander et al., 2020; Jung et al., 2021; Zdravkovic et al., 2020). Paraphrasing Barseghyan and Shaw (2017), we might

<sup>39</sup> Nicolacopoulos (1989), for example, provides a critical discussion as to whether the MSRP can be used to think about philosophical programmes. His answer is in the negative.

<sup>40</sup> One might likewise muse that the academic journal is dead – particularly given the advent of pre-print servers, micropublications, and other innovative publishing formats (see generally Dunleavy et al., 2020; Kupferschmidt, 2020). This would render the present formulation of the journal-MSRP obsolete – though it could be repurposed for novel publication formats.

<sup>41</sup> That is, the MSRP permits a contrast between journals within a finite window of time and across lengthy periods. Of course, it is not the only means of doing so. Not long after its origination, the impact factor was refashioned to capture long-term trends (Garfield, 1998). However, these efforts share the limitations of approaches described above.

<sup>42</sup> What’s more, efforts that focus solely on “bad” or “predatory” journals may be of limited value (see Eve & Priego, 2017) in the grand scheme of things – when considering that even “good” or “legitimate” journals are susceptible to polluting the scholarly corpus (e.g., Wakefield et al., 1998).

say that the MSRP of journals benefits us, over current alternatives, by providing criteria and logic for “keeping score” (p. 9) of which journals are performing best (in a given moment or timeframe), thus informing our overall assessment/appraisal and aiding in comparing multiple journals in a given field or area of inquiry.

Lastly, there is a broader set of issues that impact the development and functioning of academic journals, which deserve mention. Individual journals and publishers are vulnerable to a variety of internal and external pressures that have thus far only been alluded to. The commercialization of scholarly publishing (Larivière et al., 2015; Tennant, 2018) and research (Resnik, 2007), a skewed incentive structure for researchers (see Ritchie, 2020, Chapter 7), and (in some disciplines) the rampant influence of pharmaceutical companies (Goldacre, 2012), have shaped how science is conducted, and fostered financial and intellectual conflicts of interest among (many) authors and editorial board members (e.g., Abramson & Starfield, 2005; Brody, 2007). This has enabled the manipulation of the scholarly literature (e.g., Angell, 2000; Oreskes & Conway, 2010) through practices such as ghostwriting (McHenry, 2010; Sismondo, 2009). These issues impact not only what scholarship is published, but how it is conducted, funded, and the standards by which it is appraised and critiqued. While these issues continue to warrant greater attention, it can only be noted here that they must be taken into account when assessing the scholarly ecosystem using the MSRP framework. This will be a task required of future work – once the components of the MSRP are fully articulated, debated, and defended.

## 6 Conclusion

This article has attempted to outline some of the deficiencies of current efforts to evaluate journal quality – specifically highlighting issues with the impact factor and predatory blacklists and whitelists. It was noted that many of these efforts share features found across debates within twentieth century philosophy of science – particularly the demarcation problem. It was then proposed that Lakatos’ MSRP and his classifications of “progressive” and “degenerative” programmes could be repurposed to evaluate journals and scientific publishing more broadly. In a preliminary effort, the MI and SI were described, to help operationalize these classifications. Future theoretical and empirical work will need to be done to further explicate the characteristics of progressive and degenerative journals described here and determine the utility and applicability of these classifications for scientific and scholarly use.<sup>43</sup>

---

<sup>43</sup> Reviewer #2 has helpfully pointed out that there may be instances when such guidance is eclipsed by non-epistemic considerations. Reviewer #1 has keenly noted that it is still unclear how we may appraise journals which are obviously degenerating, yet still periodically make ground-breaking contributions.

**Acknowledgements** Thanks to Sarah Wieten, Prasanta Bandyopadhyay, Julia Dabravolskaj, Jessie Seiler, Matt DeCarlo, and Josh Nicholson for helpful discussion and comments on an earlier version of this manuscript. Members of the MetaMelb research group and the METRICS International Forum provided similarly helpful feedback. All errors and omissions are the responsibility of the above author.

## Declarations

**Conflicts of interest** The author declares that he has no conflicts of interest to disclose.

## References

- Abramson, J., & Starfield, B. (2005). The effect of conflict of interest on biomedical research and clinical practice guidelines: Can we trust the evidence in evidence-based medicine? *The Journal of the American Board of Family Practice*, 18(5), 414–418. <https://doi.org/10.3122/jabfm.18.5.414>
- Akça, S., & Akbulut, M. (2021). Are predatory journals contaminating science? An analysis on the Cabells' predatory report. *The Journal of Academic Librarianship*, 47(4), 102366. <https://doi.org/10.1016/j.acalib.2021.102366>
- Akerlof, G. A., & Michailat, P. (2019). Persistence of false paradigms in low-power sciences. *PNAS*, 115(52), 13228–13233. <https://doi.org/10.1073/pnas.1816454115>
- Alexander, P. E., Debono, V. B., Mammen, M. J., Iorio, A., Aryal, K., Deng, D., Brocard, E., & Alhazani, W. (2020). COVID-19 coronavirus research has overall low methodological quality thus far: Case in point for chloroquine/hydroxychloroquine. *Journal of Clinical Epidemiology*, 123, 120–126. <https://doi.org/10.1016/j.jclinepi.2020.04.016>
- Alfano, M. (2013). Identifying and defending the hard core of virtue ethics. *Journal of Philosophical Research*, 38, 233–260. <https://doi.org/10.5840/jpr20133812>
- Andersen, L. E., & Wray, K. B. (2019). Detecting errors that result in retractions. *Social Studies of Science*, 49(6), 942–954. <https://doi.org/10.1177/0306312719872008>
- Anderson, R. (2015). Should we retire the term “predatory publishing”? *The Scholarly Kitchen* [Blog]. Retrieved October 27, 2022, from <https://scholarlykitchen.sspnet.org/2015/05/11/should-we-retire-the-term-predatory-publishing/>
- Angell, M. (2000). Is academic medicine for sale? *New England Journal of Medicine*, 342(20), 1516–1518. <https://doi.org/10.1056/NEJM200005183422009>
- Aromataris, E., & Stern, C. (2020). Supporting a definition of predatory publishing. *BMC Medicine*, 18, 125. <https://doi.org/10.1186/s12916-020-01599-6>
- Barré-Sinoussi, F., Chermann, C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dautet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), 868–871. <https://doi.org/10.1126/science.6189183>
- Barseghyan, H., & Shaw, J. (2017). How can a taxonomy of stances help clarify classical debates on scientific change? *Philosophies*, 2(4), 24. <https://doi.org/10.3390/philosophies2040024>
- Bartley, W. W., III. (1976). On Imre Lakatos. In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Essays in memory of Imre Lakatos* (pp. 37–38). Holland. [https://doi.org/10.1007/978-94-010-1451-9\\_4](https://doi.org/10.1007/978-94-010-1451-9_4)
- Baždarić, K., Bilić-Zulle, L., Brumini, G., & Petrovečki, M. (2012). Prevalence of plagiarism in recent submissions to the Croatian Medical Journal. *Science and Engineering Ethics*, 18(2), 223–239. <https://doi.org/10.1007/s11948-011-9347-2>
- Beall, J. (2010). “Predatory” open-access scholarly publishers. *The Charleston Advisor*, 11(4), 10–17.
- Beall, J. (2014). Unintended consequences: The rise of predatory publishers and the future of scholarly publishing. *Editorial Office News*, 7(1), 4–6.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407–425. <https://doi.org/10.1037/a0021524>
- Bisaccio, M. (2018). Cabells' journal whitelist and blacklist: Intelligent data for informed journal evaluations. *Learned Publishing*, 31(3), 243–248. <https://doi.org/10.1002/leap.1164>

- Blaug, M. (1976). Kuhn versus Lakatos or paradigms versus research programmes in the history of economics. In S. Latsis (Ed.), *Method and appraisal in economics* (pp. 149–180). Cambridge University Press. <https://doi.org/10.1017/CBO9780511572203.007>
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major across multiple large samples. *American Journal of Psychiatry*, 176(5), 376–387. <https://doi.org/10.1176/appi.ajp.2018.18070881>
- Bordignon, F. (2020). Self-correction of science: A comparative study of negative citations and post-publication peer review. *Scientometrics*, 124(2), 1225–1239. <https://doi.org/10.1007/s11192-020-03536-z>
- Brembs, B. (2018). Prestigious science journals struggle to reach even average reliability. *Frontiers in Human Neuroscience*, 12, 37. <https://doi.org/10.3389/fnhum.2018.00037>
- Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7, 291. <https://doi.org/10.3389/fnhum.2013.00291>
- Brody, H. (2007). *Hooked: Ethics, the medical profession, and the pharmaceutical industry*. Rowman & Littlefield Publishers.
- Cai, N., Choi, K. W., & Fried, E. I. (2020). Reviewing the genetics of heterogeneity in depression: Operationalizations, manifestations, and etiologies. *Human Molecular Genetics*, 29(R1), R10–R18. <https://doi.org/10.1093/hmg/ddaa115>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Campbell, D. T. (1987). Evolutionary epistemology. In G. Radnitzky & W. W. Bartley (Eds.), *Evolutionary epistemology, rationality, and the sociology of knowledge* (pp. 47–89). Open Court.
- Caputo, R. K. (2019). Peer review: A vital gatekeeping function and obligation of professional scholarly practice. *Families in Society: The Journal of Contemporary Social Services*, 100(1), 6–16. <https://doi.org/10.1177/1044389418808155>
- Chalmers, A. F. (1999). *What is this thing called science?* (3rd ed.). Hackett Publishing Company, Inc.
- Chambers, C. (2017). *The 7 deadly sins of psychology: A manifesto for reforming the future of scientific practice*. Princeton University Press.
- Chen, X. (2019). Beall's list and Cabell's blacklist: A comparison of two lists of predatory OA journals. *Serials Review*, 45(4), 219–226. <https://doi.org/10.1080/00987913.2019.1694810>
- Cobey, K. D., Lalu, M. M., Skidmore, B., Ahmadzai, N., Grudniewicz, A., & Moher, D. (2018). What is a predatory journal? A scoping review. *F1000Research*, 7. <https://doi.org/10.12688/f1000research.15256.2>
- Cohen, R. S., Feyerabend, P. K., & Wartofsky, M. W. (Eds.). (1976). *Essays in memory of Imre Lakatos*. Holland. <https://doi.org/10.1007/978-94-010-1451-9>
- Correia, L. C., & Segundo, J. D. B. (2020). An immunization program against the COVID-19 infodemic. *Journal of Evidence-Based Healthcare*, 2(1), 7–9. <https://doi.org/10.17267/2675-021Xevidence.v2i1.3124>
- Crawford, W. (2014). Ethics and access 1: The sad case of Jeffrey Beall. *Cites & Insights*, 14(4), 1–14.
- Csiszar, A. (2018). *The scientific journal: Authorship and the politics of knowledge in the nineteenth century*. The University of Chicago Press.
- D'Amour, G. (1976). Research programs, rationality, and ethics. In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Essays in memory of Imre Lakatos* (pp. 87–98). Holland. <https://doi.org/10.1007/978-94-010-1451-9>
- de Marchi, N., & Blaug, M. (Eds.). (1991). *Appraising economic theories: Studies in the methodology of research programs*. Elgar.
- Demeter, M. (2020). Gatekeepers of knowledge Dissemination: Inequality in journal editorial boards. In M. Demeter (Ed.), *Academic knowledge production and the global south: Questioning inequality and under-representation* (pp. 137–151). Springer Publishing. [https://doi.org/10.1007/978-3-030-52701-3\\_6](https://doi.org/10.1007/978-3-030-52701-3_6)
- Duesberg, P. H. (1992). AIDS acquired by drug consumption and other non-contagious risk factors. *Pharmacology & Therapeutics*, 55(3), 201–277. [https://doi.org/10.1016/0163-7258\(92\)90052-2](https://doi.org/10.1016/0163-7258(92)90052-2)
- Duesberg, P. H. (Ed.). (1996a). *AIDS: Virus- or drug induced?* Kluwer Academic Publishers.
- Duesberg, P. H. (1996b). *Inventing the AIDS virus*. Regnery Publishing Inc.

- Duesberg, P. H., Koehnlein, C., & Rasnick, D. (2003). The chemical bases of the various AIDS epidemics: Recreational drugs, anti-viral chemotherapy and malnutrition. *Journal of Biosciences*, 28(4), 383–412. <https://doi.org/10.1007/bf02705115>
- Dunleavy, D. J. (2020). *Appraising contemporary social work research: Meta-research on statistical reporting, statistical power, and evidential value* [Unpublished dissertation]. Florida State University. <https://doi.org/10.31237/osf.io/f2c7b>
- Dunleavy, D. J. (2021). The cultivation of social work knowledge: Towards a more robust system of peer review. *Families in Society: The Journal of Contemporary Social Services*, 102(4), 556–568. <https://doi.org/10.31235/osf.io/rvqze>
- Dunleavy, D. J. (2022). It's time to terminate social work's relationship with the impact factor. *Social Work*, 67(3), 296–297. <https://doi.org/10.1093/sw/swac017>
- Dunleavy, D. J., & Hendricks, V. F. (2020). Fast science, slow science: Finding balance in the time of COVID-19 and the age of misinformation. Zenodo. <https://doi.org/10.5281/zenodo.4056908>
- Dunleavy, D. J., Tennant, J., & Francuzik, W. (2020). Scientific practice in the time of COVID19: It's time to commit to change. Zenodo. <https://doi.org/10.5281/zenodo.3873630>
- Dyson, F. W., Eddington, A. S., & Davidson, C. (1920). IX. A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London (A)*, 220, 291–333. <https://doi.org/10.1098/rsta.1920.0009>
- Einstein, A. (1905). Zur elektrodynamik bewegter Körper. *Annalen Der Physik*, 322(10), 891–921. <https://doi.org/10.1002/andp.19053221004>
- Elgin, M., & Sober, E. (2017). Popper's shifting appraisal of evolutionary theory. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 7, 31–55. <https://doi.org/10.1086/691119>
- Ellison, B. J., & Duesberg, P. H. (1994). *Why we will never win the war on AIDS*. Inside Story Communications.
- Ellison, B. J., Downey, A. B., & Duesberg, P. H. (1995). HIV as a surrogate marker for drug use: A reanalysis of the San Francisco Men's Health Study. *Genetica*, 95, 165–171. <https://doi.org/10.1007/BF01435008>
- Elman, C., & Elman, M. F. (2003). *Progress in international relations theory: Appraising the field*. MIT Press.
- Else, J. F. (1978). Social work journals: Purposes and trends. *Social Work*, 23(4), 267–273. <https://doi.org/10.1093/sw/23.4.267>
- Eve, M. P., & Priego, E. (2017). Who is actually harmed by predatory publishers? *tripleC*, 15(2), 755–770. <https://doi.org/10.31269/triplec.v15i2.867>
- Eyre-Walker, A., & Stoletzki, N. (2013). The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. *PLOS Biology*, 11(10), e1001675. <https://doi.org/10.1371/journal.pbio.1001675>
- Fanelli, D. (2013). Why growing retractions are (mostly) a good sign. *PLOS Medicine*, 10(12), e1001563. <https://doi.org/10.1371/journal.pmed.1001563>
- Farrell, R. P. (2003). *Feyerabend and scientific values: Tightrope-walking rationality*. Springer Publishing. <https://doi.org/10.1007/978-94-017-1542-3>
- Fernandez-Cano, A. (2021). Letter to the editors: Publish, publish ... cursed! *Scientometrics*, 126, 3673–3682. <https://doi.org/10.1007/s11192-020-03833-7>
- Feyerabend, P. (1970). Consolations for the specialist. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 197–230). Cambridge University Press.
- Feyerabend, P. (1974). Popper's objective knowledge. *Inquiry*, 17, 1–4. <https://doi.org/10.1080/00201747408601743>
- Feyerabend, P. (1975). Imre Lakatos. *The British Journal for the Philosophy of Science*, 26(1), 1–18.
- Feyerabend, P. (1980). *Against method* (3rd ed.). Verso Books.
- Frandsen, T. F. (2019). How can a questionable journal be identified: Frameworks and checklists. *Learned Publishing*, 32(3), 221–226. <https://doi.org/10.1002/leap.1230>
- Gallo, R. C. (1991). *Virus hunting: AIDS, cancer, and the human retrovirus: A story of scientific discovery*. Basic Books.
- Gallo, R. C., & Montagnier, L. (2003). The discovery of HIV as the cause of AIDS. *New England Journal of Medicine*, 349(24), 2283–2285. <https://doi.org/10.1056/nejmp038194>
- Gambrill, E. (2018). The promotion of avoidable ignorance in the *British Journal of Social Work Research on Social Work Practice*, 29(4), 455–469. <https://doi.org/10.1177/1049731518763953>

- Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479. <https://doi.org/10.1126/science.178.4060.471>
- Garfield, E. (1998). Long-term vs. short-term journal impact: Does it matter? *The Scientist*, 12(3), 10–12.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, 295(1), 90–93. <https://doi.org/10.1001/jama.295.1.90>
- Gillies, D. (2002). Lakatos' criticisms of Popper. In G. Kampis, L. Kvasz, & M. Stöltzner (Eds.) *Appraising Lakatos* (pp. 13–22). Springer Publishing. [https://doi.org/10.1007/978-94-017-0769-5\\_2](https://doi.org/10.1007/978-94-017-0769-5_2)
- Godlee, F. (2011). Wakefield's article linking MMR vaccine and autism was fraudulent. *BMJ*, 342, c7452. <https://doi.org/10.1136/bmj.c7452>
- Goldacre, B. (2008). *Bad science: Quacks, hacks, and big pharma flacks*. Fourth Estate.
- Goldacre, B. (2012). *Bad pharma: How drug companies mislead doctors and harm patients*. Fourth Estate.
- Groves, T. (2016). Lakatos's criticism of Carnapian inductive logic was mistaken. *Journal of Applied Logic*, 14, 3–21. <https://doi.org/10.1016/j.jal.2015.09.014>
- Grudniewicz, A., Moher, D., Cobey, K. D., Bryson, G. L., Cukier, S., Allen, K., Ardern, C., Balcom, L., Barros, T., Berger, M., Ciro, J. B., Cugusi, L., Donaldson, M. R., Egger, M., Graham, I. D., Hodgkinson, M., Khan, K. M., Mabizela, M., Manca, A., et al. (2019). Predatory journals: No definition, no defence. *Nature*, 576(7786), 210–212. <https://doi.org/10.1038/d41586-019-03759-y>
- Guthrie, S., Ghiga, I., & Wooding, S. (2017). What do we know about grant peer review in the health sciences? *F1000Research*, 6, 1335. <https://doi.org/10.12688/f1000research.11917.2>
- Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge University Press.
- Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. A. (2020). Calibrating the scientific ecosystem through meta-research. *Annual Review of Statistics and Its Application*, 7, 11–37. <https://doi.org/10.1146/annurev-statistics-031219-041104>
- Harvey, L. (2020). Research fraud: A long-term problem exacerbated by the clamour for research grants. *Quality in Higher Education*, 26(3), 243–261. <https://doi.org/10.1080/13538322.2020.1820126>
- Healy, D. (2015). Serotonin and depression: The marketing of a myth. *BMJ*, 350. <https://doi.org/10.1136/bmj.h1771>
- Hefner, P. (1993). *The human factor: Evolution, culture, and religion*. Augsburg Fortress.
- Hoeffel, C. (1998). Journal impact factors. *Allergy*, 53(12), 1225. <https://doi.org/10.1111/j.1398-9995.1998.tb03848.x>
- Horrobin, D. F. (1990). The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association*, 263(10), 1438–1441. <https://doi.org/10.1001/jama.1990.03440100162024>
- Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), 1262–1278. <https://doi.org/10.1016/j.cell.2014.05.010>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A. (2013). Implausible results in human nutrition research. *BMJ*, 347, f6698. <https://doi.org/10.1136/bmj.f6698>
- Ioannidis, J. P. A. (2016). Evidence-based medicine has been hijacked: A report to David Sackett. *Journal of Clinical Epidemiology*, 73, 82–86. <https://doi.org/10.1016/j.jclinepi.2016.02.012>
- Ioannidis, J. P. A., & Thombs, B. D. (2019). A user's guide to inflated and manipulated impact factors. *European Journal of Clinical Investigation*, 49(9), e13151. <https://doi.org/10.1111/eci.13151>
- JAMA: Journal of the American Medical Association. (2021). *Instructions for authors*. Retrieved October 27, 2022, from <https://jamanetwork.com/journals/jama/pages/instructions-for-authors>
- Johnson, R., Watkinson, A., & Mabe, M. (2018). *The STM report: An overview of scientific and scholarly publishing* (5th ed.). Retrieved October 27, 2022, from [https://www.stm-assoc.org/2018\\_10\\_04\\_STM\\_Report\\_2018.pdf](https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf)
- Jung, R. G., Santo, P. D., Clifford, C., Prospero-Porta, G., Skanes, S., Hung, A., Parlow, S., Visintini, S., Ramirez, F. D., Simard, T., & Hibbert, B. (2021). Methodological quality of COVID-19 clinical research. *Nature Communications*, 12, 943. <https://doi.org/10.1038/s41467-021-21220-5>
- Kadvany, J. (2001). *Imre Lakatos and the guises of reason*. Duke University Press.
- Kendall, G. (2021). Beall's legacy in the battle against predatory publishers. *Learned Publishing*, 34(3), 379–388. <https://doi.org/10.1002/leap.1374>

- Koertge, N. (1972). For and against method. *The British Journal for the Philosophy of Science*, 23(3), 274–290. <https://doi.org/10.1093/bjps/23.3.274>
- Kratochvíl, J., Plch, L., Sebera, M., & Koriřáková, E. (2020). Evaluation of untrustworthy journals: Transition from formal criteria to a complex view. *Learned Publishing*, 33(3), 308–322. <https://doi.org/10.1002/leap.1299>
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. The University of Chicago Press.
- Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 1–23). Cambridge University Press.
- Kuhn, T. S. (1980). The halt and the blind: Philosophy and history of science. *The British Journal for the Philosophy of Science*, 31(2), 181–192.
- Kupferschmidt, K. (2020). Preprints bring ‘firehose’ of outbreak data. *Science*, 367(6481), 963–964. <https://doi.org/10.1126/science.367.6481.963>
- Lacasse, J. R., & Leo, J. (2005). Serotonin and depression: A disconnect between the advertisements and the scientific literature. *PLOS Medicine*, 2(12), e392, 1211–1216. <https://doi.org/10.1371/journal.pmed.0020392>
- Laine, C., & Winker, M. A. (2017). Identifying predatory or pseudo-journals. *Biochemia Medica*, 27(2), 285–291. <https://doi.org/10.11613/BM.2017.031>
- Lakatos, I. (1968a). Changes in the problem of inductive logic. In I. Lakatos (Ed.), *The problem of inductive logic* (pp. 315–417). North-Holland Publishing Company.
- Lakatos, I. (1968b). Criticism and the methodology of scientific research programmes. *Proceedings of the Aristotelian Society*, 69, 149–186.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.
- Lakatos, I. (1971a). The history of science and its rational reconstructions. In R.C. Buck & R.S. Cohen (Eds.), *PSA: Proceedings of the Biennial Meetings of the Philosophy of Science Association* (pp. 91–139). Springer Publishing. [https://doi.org/10.1007/978-94-010-3142-4\\_7](https://doi.org/10.1007/978-94-010-3142-4_7)
- Lakatos, I. (1971b). Replies to critics. In R. C. Buck & R. S. Cohen (Eds.), *PSA: Proceedings of the Biennial Meetings of the Philosophy of Science Association* (pp. 174–182). Springer Publishing. [https://doi.org/10.1007/978-94-010-3142-4\\_12](https://doi.org/10.1007/978-94-010-3142-4_12)
- Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge University Press.
- Lakatos, I. (1978a). *Mathematics, science, and epistemology*. Cambridge University Press.
- Lakatos, I. (1978b). *The methodology of scientific research programmes*. Cambridge University Press.
- Lakatos, I. (2012). Science and pseudoscience. In M. Curd, J. A. Cover, & C. Pincock (Eds.), *Philosophy of science: The central issues* (2nd ed., pp. 20–26). W. W. Norton.
- Lakatos, I., & Zahar, E. (1975). Why did Copernicus’ programme supersede Ptolemy’s? In R. Westman (Ed.), *The Copernican achievement* (pp. 354–383). University of California Press.
- Lakatos, I., & Musgrave, A. (Eds.). (1970). *Criticism and the growth of knowledge*. Cambridge University Press.
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE*, 10(6), e0127502. <https://doi.org/10.1371/journal.pone.0127502>
- Larivière, C., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Springer Publishing. [https://doi.org/10.1007/978-3-030-02511-3\\_1](https://doi.org/10.1007/978-3-030-02511-3_1)
- Larvor, B. (1998). *Lakatos: An introduction*. Routledge.
- Larvor, B. (2006). Research programs. In S. Sarkar & J. Pfeifer (Eds.), *Philosophy of science: An encyclopedia* (pp. 713–715). Routledge.
- Latsis, S. (Ed.). (1976). *Method and appraisal in economics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511572203>
- Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Routledge and Kegan Paul.
- Lock, S. (1985). *A difficult balance: Editorial peer review in medicine*. ISI Press.
- Lock, S. (1989). “Journalology”: Are the quotes needed. *CBE Views*, 12(4), 57–59.
- Lorrimar, V. (2017). Are scientific research programmes applicable to theology? On Philip Hefner’s use of Lakatos. *Theology and Science*, 15(2), 188–202. <https://doi.org/10.1080/14746700.2017.1299376>

- Margalida, A., & Colomer, M. À. (2015). Mistake index as a surrogate of quality in scientific manuscripts. *Proceedings of the National Academy of Sciences*, 112(13), E1511. <https://doi.org/10.1073/pnas.1500322112>
- Margalida, A., & Colomer, M. À. (2016). Improving the peer-review process and editorial quality: Key errors escaping the review and editorial process in top scientific journals. *PeerJ*, 4, e1670. <https://doi.org/10.7717/peerj.1670>
- Mayo, D. G. (2018). *Statistical inference as severe testing*. Cambridge University Press.
- McCloskey, D. N. (1993). Review of "Appraising economic theories: Studies in the methodology of research programs" Edited by Neil de Marchi and Mark Blaug. *Journal of Economic Literature*, 31(1), 229–231.
- McHenry, L. (2010). Of sophists and spin-doctors: Industry-sponsored ghostwriting and the crisis of academic medicine. *Mens Sana Monographs*, 8(1), 129–145. <https://doi.org/10.4103/0973-1229.58824>
- McLaughlin, K. A., & Hatzenbuehler, M. L. (2009). Mechanisms linking stressful life events and mental health problems in a prospective, community-based sample of adolescents. *Journal of Adolescent Health*, 44(2), 153–160. <https://doi.org/10.1016/j.jadohealth.2008.06.019>
- McNutt, M. (2016). Taking up TOP. *Science*, 352(6290), 1147. <https://doi.org/10.1126/science.aag2359>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. The University of Chicago Press.
- Mollett, A., Brumley, C., Gilson, C., & Williams, S. (2017). *Communicating your research with social media: A practical guide to using blogs, podcasts, data visualisations and video*. Sage.
- Motterlini, M. (1995). Has Lakatos really gone a long way towards epistemological anarchism? *Epistemologia*, XVIII, 215–232.
- Motterlini, M. (Ed.). (1999). *For and against method: Including Lakatos's lectures on scientific method and the Lakatos-Feyerabend correspondence*. University of Chicago Press.
- Motterlini, M. (2002). Professor Lakatos between the Hegelian devil and the Popperian deep blue sea. In G. Kampis, L. Kvasz, & M. Stöltzner (Eds.), *Appraising Lakatos* (pp. 23–52). Springer Publishing. [https://doi.org/10.1007/978-94-017-0769-5\\_3](https://doi.org/10.1007/978-94-017-0769-5_3)
- Murphy, N. (1990). *Theology in the age of scientific reasoning*. Cornell University Press.
- Musgrave, A. (1976a). Method or madness? In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Essays in memory of Imre Lakatos* (pp. 457–491). Holland. [https://doi.org/10.1007/978-94-010-1451-9\\_27](https://doi.org/10.1007/978-94-010-1451-9_27)
- Musgrave, A. (1976b). Why did oxygen supplant phlogiston? Research programmes in the Chemical Revolution. In C. Howson (Ed.), *Method and appraisal in the physical sciences: The Critical background to modern science, 1800–1905* (pp. 181–210). Cambridge University Press. <https://doi.org/10.1017/CBO9780511760013.005>
- Musgrave, A. (1978). Evidential support, falsification, heuristics, and anarchism. In G. Radnitzky & G. Andersson (Eds.), *Progress and rationality in science* (pp. 181–201). Springer Publishing. [https://doi.org/10.1007/978-94-009-9866-7\\_8](https://doi.org/10.1007/978-94-009-9866-7_8)
- Musgrave, A. & Pigden, C. (2016). Imre Lakatos. *Stanford Encyclopedia of Philosophy*. Retrieved October 27, 2022 from <https://plato.stanford.edu/entries/lakatos/>
- Nature. (2005). Not-so-deep impact. *Nature*, 435(7045), 1003–1004. <https://doi.org/10.1038/4351003b>
- New England Journal of Medicine. (n.d.). *About NEJM*. Retrieved October 27, 2022, from <https://www.nejm.org/about-nejm/about-nejm>
- Neylon, C. (2017). *Blacklists are technically infeasible, practically unreliable and unethical*. Period [Blog]. LSE Impact Blog. Retrieved October 27, 2022, from <http://eprints.lse.ac.uk/70018/1/blogs.lse.ac.uk-Blacklists%20are%20technically%20infeasible%20practically%20unreliable%20and%20unethical%20Period.pdf>
- Ng, J. Y., & Haynes, R. B. (2021). "Evidence-based checklists" for identifying predatory journals have not been assessed for reliability or validity: An analysis and proposal for moving forward. *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2021.06.015>
- Nicholson, J. M. (2021). Smart(er) citations. *Matter*, 4(3), 756–758. <https://doi.org/10.1016/j.matt.2021.02.007>
- Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rorigues, N. P., Grabitz, P., & Rife, S. C. (2021). *Scite: A smart citation index that displays the context of citations and classifies their intent using deep learning*. bioRxiv. <https://doi.org/10.1101/2021.03.15.435418>
- Nicolacopoulos, P. D. (1989). Through the looking glass: Philosophy, research programmes and the scientific community. In K. Gavroglu, T. Goudaroulis, & P. Nicolacopoulos (Eds.), *Imre Lakatos and theories of*



- scientific change* (pp. 189–202). Kluwer Academic Publishers. [https://doi.org/10.1007/978-94-009-3025-4\\_15](https://doi.org/10.1007/978-94-009-3025-4_15)
- Nosek, B. A., Spies, J. R., & Nytly, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- O'Connor, C. (2019). The natural selection of conservative science. *Studies in History and Philosophy of Science A*, 76, 24–29. <https://doi.org/10.1016/j.shpsa.2018.09.007>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 943–951. <https://doi.org/10.1126/science.aac4716>
- Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.
- Phillips, D. C. (1987). *Philosophy, science, and social inquiry: Contemporary methodological controversies in social science and related applied fields of research*. Pergamon Press.
- Phillips, D. C., & Nicolayev, J. (1978). Kohlbergian moral development: A progressing or degenerating research program? *Educational Theory*, 28(4), 286–301. <https://doi.org/10.1111/j.1741-5446.1978.tb00824.x>
- PLOS One. (2019). *Data availability*. Retrieved October 27, 2022, from <https://journals.plos.org/plosone/data-availability>
- Popper, K. R. (1959/1968). *The logic of scientific discovery*. Harper & Row Publishers.
- Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge & Kegan Paul plc.
- Popper, K. R. (1970). Normal science and its dangers. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 51–58). Cambridge University Press.
- Popper, K. R. (1974). Intellectual autobiography. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 3–181). Open Court.
- Popper, K. R. (1982). *Quantum theory and the schism in physics: From the Postscript to the Logic of Scientific Discovery*. Routledge.
- Porter, R. (1992). The rise of medical journalism in Britain to 1800. In W. F. Bynam, S. Lock, & R. Porter (Eds.), *Medical journals and medical knowledge: Historical essays* (pp. 6–28). Routledge.
- Psychological Science. (n.d.). *Journal description*. Retrieved October 27, 2022, from <https://journals.sagepub.com/description/PSS>
- Quintana, D. S., & Heathers, J. A. (2021). How podcasts can benefit scientific communities. *Trends in Cognitive Sciences*, 25(1), 3–5. <https://doi.org/10.1016/j.tics.2020.10.003>
- Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge*. The University of Chicago Press.
- Resnik, D. (2007). *The price of truth: How money affects the norms of science*. Oxford University Press.
- Ritchie, S. (2020). *Science fictions: How fraud, bias, negligence, and hype undermine the search for truth*. Metropolitan Books.
- Rossner, M., Epps, H. V., & Hill, E. (2007). Show me the data. *The Journal of Cell Biology*, 179(6), 1091–1092. <https://doi.org/10.1083/jcb.200711140>
- Rothman, K. J. (1998). Writing for Epidemiology. *Epidemiology*, 9(3), 333–337. <https://doi.org/10.1097/00001648-199805000-00019>
- Scheffler, I. (1963). *The anatomy of inquiry: Philosophical studies in the theory of science*. Knopf.
- scite. (2019a). *Journals differ in the incidence of supporting and contradicting citations they receive*. Medium. Retrieved October 27, 2022, from <https://medium.com/scite/journals-differ-in-the-incidence-of-supporting-and-contradicting-citations-they-receive-7410f08dd16b>
- scite. (2019b). *scite: Why, How, and Who*. Medium. Retrieved October 27, 2022, from <https://medium.com/scite/scite-why-how-and-who-c064f87d1a00>
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314(7079), 497. <https://doi.org/10.1136/bmj.314.7079.497>
- Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, 7(21), eabd1705. <https://doi.org/10.1126/sciadv.abd1705>
- Shamseer, L., & Moher, D. (2017). Thirteen ways to spot a 'predatory journal' (and why we shouldn't call them that). *Times Higher Education*. Retrieved October 27, 2022, from <https://www.timeshighereducation.com/blog/thirteen-ways-to-spot-a-predatory-journal-and-why-we-shouldnt-call-them-that>
- Shamseer, L., Moher, D., Maduekwe, O., Turner, L., Barbour, V., Burch, R., Clark, J., Galipeau, J., Roberts, J., & Shea, B. J. (2017). Potential predatory and legitimate biomedical journals: Can you tell

- the difference? A cross-sectional comparison. *BMC Medicine*, 15, 28. <https://doi.org/10.1186/s12916-017-0785-9>
- Siler, K. (2020). Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy. *Journal of the Association for Information Science and Technology*, 71(11), 1386–1401. <https://doi.org/10.1002/asi.24339>
- Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, 112(2), 360–365. <https://doi.org/10.1073/pnas.1418218112>
- Simmons, J., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Sismondo, S. (2009). Ghosts in the machine: Publication planning in the medical sciences. *Social Studies of Science*, 39(2), 171–198. <https://doi.org/10.1177/0306312708101047>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. <https://doi.org/10.1098/rsos.160384>
- Smith, R. (2006). The trouble with medical journals. *Journal of the Royal Society of Medicine*, 99, 115–119. <https://doi.org/10.1258/jrsm.99.3.115>
- Social Work Research. (n.d.). *About the journal*. Retrieved October 27, 2022, from <https://academic.oup.com/swr/pages/About>
- Sternberg, R. J. (2018). Evaluating merit among scientists. *Journal of Applied Research in Memory and Cognition*, 7(2), 209–216. <https://doi.org/10.1016/j.jarmac.2018.03.003>
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691612460687>
- Taubes, G. (2007). *Good calories, bad calories: Fats, carbs, and the controversial science of diet and health*. Anchor Books.
- Teixeira da Silva, J. A., & Tsigaris, P. (2018). What value do journal whitelists and blacklists have in academia?. *The Journal of Academic Librarianship*, 44(6), 781–792. <https://doi.org/10.1016/j.acalib.2018.09.017>
- Teixeira da Silva, J. A., Dobránszki, J., Tsigaris, P., & Al-Khatib, A. (2019). Predatory and exploitative behaviour in academic publishing: An assessment. *The Journal of Academic Librarianship*, 45(6), 102071. <https://doi.org/10.1016/j.acalib.2019.102071>
- Teixeira da Silva, J. A., Dunleavy, D. J., Moradzadeh, M., & Eykens, J. (2021). A credit-like rating system to determine the legitimacy of scientific journals and publishers. *Scientometrics*, 126, 8589–8616. <https://doi.org/10.1007/s11192-021-04118-3>
- Tennant, J. (2018). *Democratising knowledge: A report on the scholarly publisher, Elsevier*. Education International. Retrieved October 27, 2022, from <https://esac-initiative.org/democratising-knowledge-a-report-on-the-scholarly-publisher-elsevier-by-dr-jonathan-tennant/>
- Tennant, J. (2020). *The [r]evolution of open science*. Zenodo. <https://doi.org/10.5281/zenodo.3700646>
- Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatib, Y., B. Collister, L., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheim, S., Rigling, L., Katz, D. S., Greshake Tzovaras, B., et al. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, 6, 1151. <https://doi.org/10.12688/f1000research.12037.3>
- Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B. S., Ross-Hellauer, T., Sattler, S., Thacker, P., & Vanholsbeeck, M. (2019). Ten myths around open scholarly publishing (e27580v1). *PeerJ*. <https://doi.org/10.7287/peerj.preprints.27580v1>
- Tobin, M. J. (2004). Assessing the performance of a medical journal. *American Journal of Respiratory and Critical Care Medicine*, 169(12), 1268–1272. <https://doi.org/10.1164/rccm.2404006>
- Trafimow, D. (2014). Editorial. *Basic and Applied Social Psychology*, 36, 1–2. <https://doi.org/10.1080/01973533.2014.865505>
- Tsigaris, P., & Teixeira da Silva, J. (2021). Why blacklists are not reliable: A theoretical framework. *The Journal of Academic Librarianship*, 47(1), 102266. <https://doi.org/10.1016/j.acalib.2020.102266>
- Urbach, P. (1974a). Progress and degeneration in the ‘IQ debate’ (I). *The British Journal for the Philosophy of Science*, 25(2), 99–135.
- Urbach, P. (1974b). Progress and degeneration in the ‘IQ debate’ (II). *The British Journal for the Philosophy of Science*, 25(3), 235–259.
- Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., van Os, J., & Bentall, R. P. (2012). Childhood adversities increase the risk of psychosis: A meta-analysis of

- patient-control, prospective- and cross-sectional cohort studies. *Schizophrenia Bulletin*, 38(4), 661–671. <https://doi.org/10.1093/schbul/sbs050>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., Berelowitz, M., Dhillon, A. P., Thomson, M. A., Harvey, P., Valentine, A., Davies, S. E., & Walker-Smith, J. A. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*, 351(9103), 637–641. [https://doi.org/10.1016/s0140-6736\(97\)11096-0](https://doi.org/10.1016/s0140-6736(97)11096-0)
- Wickham, S., Taylor, P., Shevlin, M., & Bentall, R. P. (2014). The impact of social deprivation on paranoia, hallucinations, mania and depression: The role of discrimination social support, stress and trust. *PLoS ONE*, 9(8), e105140. <https://doi.org/10.1371/journal.pone.0105140>
- Worrall, J. (1978). The ways in which the methodology of scientific research programmes improves on Popper's methodology. In G. Radnitzky & G. Andersson (Eds.), *Progress and rationality in science* (pp. 45–70). Springer Publishing. [https://doi.org/10.1007/978-94-009-9866-7\\_3](https://doi.org/10.1007/978-94-009-9866-7_3)
- Wray, K. B., & Andersen, L. E. (2018). Retractions in Science. *Scientometrics*, 117, 2009–2019. <https://doi.org/10.1007/s11192-018-2922-4>
- Zahar, E. (1973a). Why did Einstein's programme supersede Lorentz's? (I). *The British Journal for the Philosophy of Science*, 24(2), 95–123.
- Zahar, E. (1973b). Why did Einstein's programme supersede Lorentz's? (II). *The British Journal for the Philosophy of Science*, 24(3), 223–262.
- Zdravkovic, M., Berger-Estilita, J., Zdravkovic, B., & Berger, D. (2020). Scientific quality of COVID-19 and SARS CoV-2 publications in the highest impact medical journals during the early phase of the pandemic: A case control study. *PLoS ONE*, 15(11), e0241826. <https://doi.org/10.1371/journal.pone.0241826>
- Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, 9(1), 66–100. <https://doi.org/10.1007/BF01553188>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.