Deep impact: unintended consequences of journal rank

Björn Brembs[1]*, Katherine Button[2] and Marcus Munafò[3]

[1] Institute of Zoology—Neurogenetics, University of Regensburg, Regensburg, Germany
[2] School of Social and Community Medicine, University of Bristol, Bristol, UK
[3] UK Centre for Tobacco Control Studies and School of Experimental Psychology, University of Bristol, Bristol, UK

Most researchers acknowledge an intrinsic hierarchy in the scholarly journals ("journal rank") that they submit their work to, and adjust not only their submission but also their reading strategies accordingly. On the other hand, much has been written about the negative effects of institutionalizing journal rank as an impact measure. So far, contributions to the debate concerning the limitations of journal rank as a scientific impact assessment tool have either lacked data, or relied on only a few studies. In this review, we present the most recent and pertinent data on the consequences of our current scholarly communication system with respect to various measures of scientific quality (such as utility/citations, methodological soundness, expert ratings or retractions). These data corroborate previous hypotheses: using journal rank as an assessment tool is bad scientific practice. Moreover, the data lead us to argue that any journal rank (not only the currently-favored Impact Factor) would have this negative impact. Therefore, we suggest that abandoning journals altogether, in favor of a library-based scholarly communication system, will ultimately be necessary. This new system will use modern information technology to vastly improve the filter, sort and discovery functions of the current journal system.

Keywords: impact factor, journal ranking, statistics as topic, publishing, open access, scholarly communication, libraries, library services

INTRODUCTION

Science is the bedrock of modern society, improving our lives through advances in medicine, communication, transportation, forensics, entertainment and countless other areas. Moreover, today's global problems cannot be solved without scientific input and understanding. The more our society relies on science, and the more our population becomes scientifically literate, the more important the reliability [i.e., veracity and integrity, or, "credibility" (Ioannidis, 2012)] of scientific research becomes. Scientific research is largely a public endeavor, requiring public trust. Therefore, it is critical that public trust in science remains high. In other words, the reliability of science is not only a societal imperative, it is also vital to the scientific community itself. However, every scientific publication may in principle report results which prove to be unreliable, either unintentionally, in the case of honest error or statistical variability, or intentionally in the case of misconduct or fraud. Even under ideal circumstances, science can never provide us with absolute truth. In Karl Popper's words: "Science is not a system of certain, or established, statements" (Popper, 1995). Peer-review is one of the mechanisms which have evolved to increase the reliability of the scientific literature.
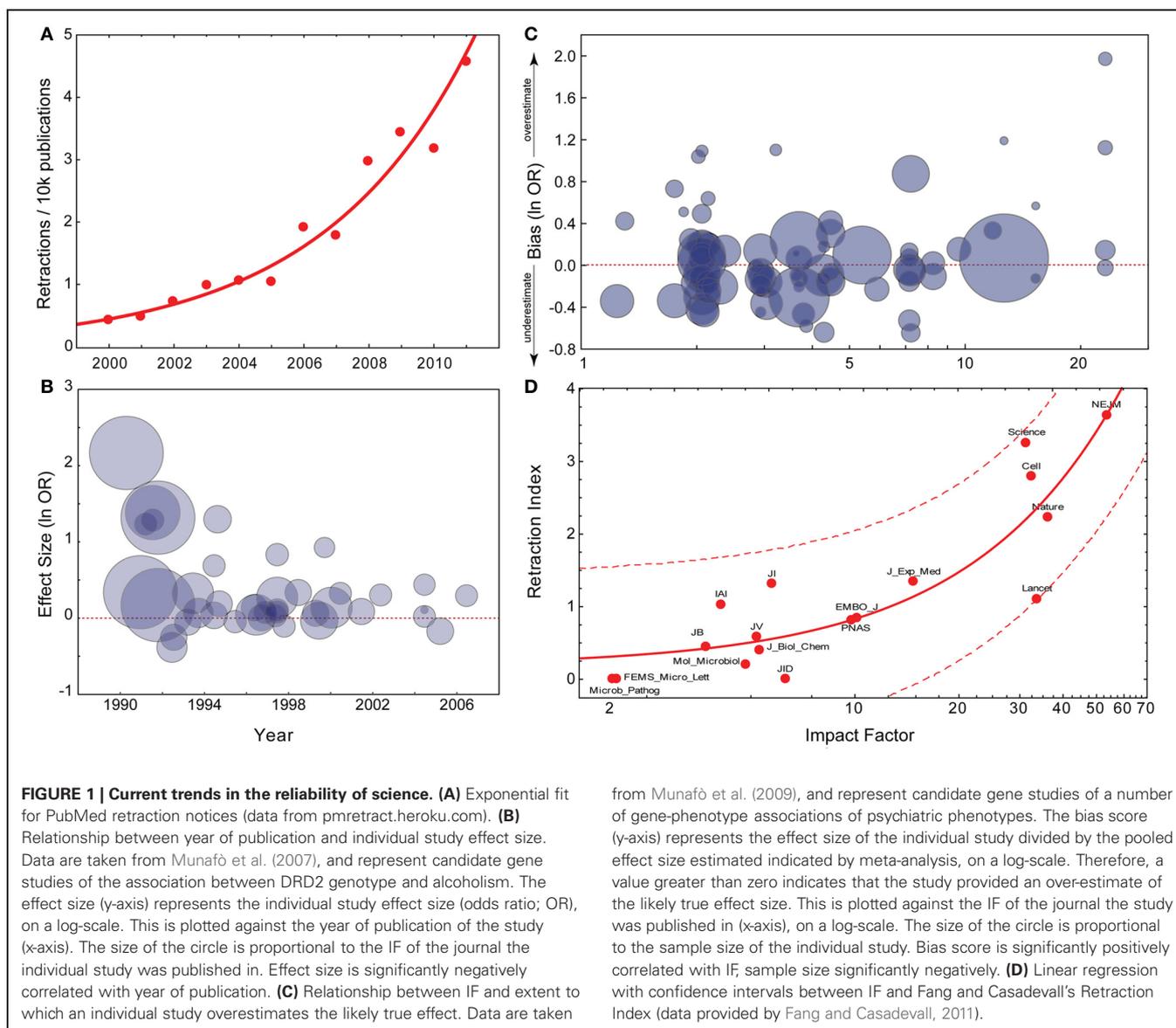
At the same time, the current publication system is being used to structure the careers of the members of the scientific community by evaluating their success in obtaining publications in high-ranking journals. The hierarchical publication system ("journal rank") used to communicate scientific results is thus central, not only to the composition of the scientific community

at large (by selecting its members), but also to science's position in society. In recent years, the scientific study of the effectiveness of such measures of quality control has grown.

RETRACTIONS AND THE DECLINE EFFECT

A disturbing trend has recently gained wide public attention: The retraction rate of articles published in scientific journals, which had remained stable since the 1970's, began to increase rapidly in the early 2000's from 0.001% of the total to about 0.02% (**Figure 1A**). In 2010 we have seen the creation and popularization of a website dedicated to monitoring retractions (http://retractionwatch.com), while 2011 has been described as the "the year of the retraction" (Hamilton, 2011). The reasons suggested for retractions vary widely, with the recent sharp rise potentially facilitated by an increased willingness of journals to issue retractions, or increased scrutiny and error-detection from online media. Although cases of clear scientific misconduct initially constituted a minority of cases (Nath et al., 2006; Cokol et al., 2007; Fanelli, 2009; Steen, 2011a; Van Noorden, 2011; Wager and Williams, 2011), the fraction of retractions due to misconduct has risen sharper than the overall retraction rate and now the majority of all retractions is due to misconduct (Steen, 2011b; Fang et al., 2012).

Retraction notices, a metric which is relatively easy to collect, only constitute the extreme end of a spectrum of unreliability that is inherent to the scientific method: we can hardly ever be entirely certain of our results (Popper, 1995). Much of the training scientists receive aims to reduce this uncertainty long before the work

FIGURE 1 | Current trends in the reliability of science. (A) Exponential fit for PubMed retraction notices (data from pmretract.heroku.com). **(B)** Relationship between year of publication and individual study effect size. Data are taken from Munafò et al. (2007), and represent candidate gene studies of the association between DRD2 genotype and alcoholism. The effect size (y-axis) represents the individual study effect size (odds ratio; OR), on a log-scale. This is plotted against the year of publication of the study (x-axis). The size of the circle is proportional to the IF of the journal the individual study was published in. Effect size is significantly negatively correlated with year of publication. **(C)** Relationship between IF and extent to which an individual study overestimates the likely true effect. Data are taken

from Munafò et al. (2009), and represent candidate gene studies of a number of gene-phenotype associations of psychiatric phenotypes. The bias score (y-axis) represents the effect size of the individual study divided by the pooled effect size estimated indicated by meta-analysis, on a log-scale. Therefore, a value greater than zero indicates that the study provided an over-estimate of the likely true effect. This is plotted against the IF of the journal the study was published in (x-axis), on a log-scale. The size of the circle is proportional to the sample size of the individual study. Bias score is significantly positively correlated with IF, sample size significantly negatively. **(D)** Linear regression with confidence intervals between IF and Fang and Casadevall's Retraction Index (data provided by Fang and Casadevall, 2011).

is submitted for publication. However, a less readily quantified but more frequent phenomenon (compared to rare retractions) has recently garnered attention, which calls into question the effectiveness of this training. The "decline-effect," which is now well-described, relates to the observation that the strength of evidence for a particular finding often declines over time (Simmons et al., 1999; Palmer, 2000; Møller and Jennions, 2001; Ioannidis, 2005b; Møller et al., 2005; Fanelli, 2010; Lehrer, 2010; Schooler, 2011; Simmons et al., 2011; Van Dongen, 2011; Bertamini and Munafo, 2012; Gonon et al., 2012). This effect provides wider scope for assessing the unreliability of scientific research than retractions alone, and allows for more general conclusions to be drawn.

Researchers make choices about data collection and analysis which increase the chance of false-positives (i.e., researcher bias) (Simmons et al., 1999; Simmons et al., 2011), and surprising and novel effects are more likely to be published than

studies showing no effect. This is the well-known phenomenon of publication bias (Song et al., 1999; Møller and Jennions, 2001; Callaham, 2002; Møller et al., 2005; Munafò et al., 2007; Dwan et al., 2008; Young et al., 2008; Schooler, 2011; Van Dongen, 2011). In other words, the probability of getting a paper published might be biased toward larger initial effect sizes, which are revealed by later studies to be not so large (or even absent entirely), leading to the decline effect. While sound methodology can help reduce researcher bias (Simmons et al., 1999), publication bias is more difficult to address. Some journals are devoted to publishing null results, or have sections devoted to these, but coverage is uneven across disciplines and often these are not particularly high-ranking or well-read (Schooler, 2011; Nosek et al., 2012). Publication therein is typically not a cause for excitement (Giner-Sorolla, 2012; Nosek et al., 2012), leading to an overall low frequency of replication studies in many fields (Kelly, 2006; Carpenter, 2012; Hartshorne and Schachner, 2012; Makel et al.,

2012; Yong, 2012). Publication bias is also exacerbated by a tendency for journals to be less likely to publish replication studies (or, worse still, failures to replicate) (Curry, 2009; Goldacre, 2011; Sutton, 2011; Editorial, 2012; Hartshorne and Schachner, 2012; Yong, 2012). Here we argue that the counter-measures proposed to improve the reliability and veracity of science such as peer-review in a hierarchy of journals or methodological training of scientists may not be sufficient.

While there is growing concern regarding the increasing rate of retractions in particular, and the unreliability of scientific findings in general, little consideration has been given to the infrastructure by which scientists not only communicate their findings but also evaluate each other as a potential contributing factor. That is, to what extent does the environment in which science takes place contribute to the problems described above? By far the most common metric by which publications are evaluated, at least initially, is the perceived prestige or rank of the journal in which they appear. Does the pressure to publish in prestigious, high-ranking journals contribute to the unreliability of science?

THE DECLINE EFFECT AND JOURNAL RANK

The common pattern seen where the decline effect has been documented is one of an initial publication in a high-ranking journal, followed by attempts at replication in lower-ranked journals which either failed to replicate the original findings, or suggested a much weaker effect (Lehrer, 2010). Journal rank is most commonly assessed using Thomson Reuters' Impact Factor (IF), which has been shown to correspond well with subjective ratings of journal quality and rank (Gordon, 1982; Saha et al., 2003; Yue et al., 2007; Sønderstrup-Andersen and Sønderstrup-Andersen, 2008). One particular case (Munafò et al., 2007) illustrates the decline effect (**Figure 1B**), and shows that early publications both report a larger effect than subsequent studies, and are also published in journals with a higher IF. These observations raise the more general question of whether research published in high-ranking journals is inherently less reliable than research in lower-ranking journals.

As journal rank is also predictive of the incidence of fraud and misconduct in retracted publications, as opposed to other reasons for retraction (Steen, 2011a), it is not surprising that higher ranking journals are also more likely to publish fraudulent work than lower ranking journals (Fang et al., 2012). These data, however, cover only the small fraction of publications that have been retracted. More important is the large body of the literature that is not retracted and thus actively being used by the scientific community. There is evidence that unreliability is higher in high-ranking journals as well, also for non-retracted publications: A meta-analysis of genetic association studies provides evidence that the extent to which a study over-estimates the likely true effect size is positively correlated with the IF of the journal in which it is published (**Figure 1C**) (Munafò et al., 2009). Similar effects have been reported in the context of other research fields (Ioannidis, 2005a; Ioannidis and Panagiotou, 2011; Siontis et al., 2011).

There are additional measures of scientific quality and in none does journal rank fare much better. A study in crystallography reports that the quality of the protein structures described

is significantly lower in publications in high-ranking journals (Brown and Ramaswamy, 2007). Adherence to basic principles of sound scientific (e.g., the CONSORT guidelines: http://www.consort-statement.org), or statistical methodology have also been tested. Four different studies on levels of evidence in medical and/or psychological research have found varying results. While two studies on surgery journals found a correlation between IF and the levels of evidence defined in the respective studies (Obremskey et al., 2005; Lau and Samman, 2007), a study of anesthesia journals failed to find any statistically significant correlation between journal rank and evidence-based medicine principles (Bain and Myles, 2005) and a study of seven medical/psychological journals found highly varying adherence to statistical guidelines, irrespective of journal rank (Tressoldi et al., 2013). The two surgery studies covered an IF range between 0.5 and 2.0, and 0.7 and 1.2, respectively, while the anesthesia study covered the range 0.8–3.5. It is possible that any correlation at the lower end of the scale is abolished when higher rank journals are included. The study by Tressoldi and colleagues, which included very high ranking journals, supports this interpretation. Importantly, if publications in higher ranking journals were methodologically sounder, then one would expect the opposite result: inclusion of high-ranking journals should result in a stronger, not a weaker correlation. Further supporting the notion that journal rank is a poor predictor of statistical soundness is our own analysis of data on statistical power in neuroscience studies (Button et al., 2013). There was no significant correlation between statistical power and journal rank (N = 650; $r_s = -0.01$; $t = 0.8$; **Figure 2**). Thus, the currently available data seem to indicate that journal rank is a poor indicator of methodological soundness.

Beyond explicit quality metrics and sound methodology, reproducibility is at the core of the scientific method and thus
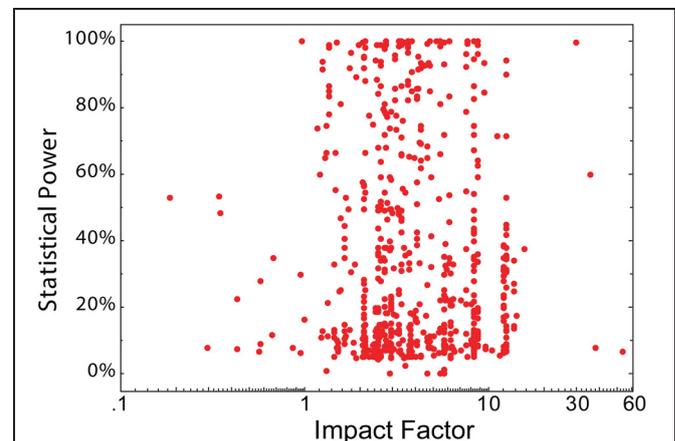

FIGURE 2 | No association between statistical power and journal IF. The statistical power of 650 neuroscience studies (data from Button et al., 2013; 19 missing ref; 3 unclear reporting; 57 published in journal without 2011 IF; 1 book) plotted as a function of the 2011 IF of the publishing journal. The studies were selected from the 730 contributing to the meta-analyses included in Button et al. (2013), **Table 1**, and included where journal title and IF (2011 © Thomson Reuters Journal Citation Reports) were available.

a hallmark of scientific quality. Three recent studies reported attempts to replicate published findings in preclinical medicine (Scott et al., 2008; Prinz et al., 2011; Begley and Ellis, 2012). All three found a very low frequency of replication, suggesting that maybe only one out of five preclinical findings is reproducible. In fact, the level of reproducibility was so low that no relationship between journal rank and reproducibility could be detected. Hence, these data support the necessity of recent efforts such as the "Reproducibility Initiative" (Baker, 2012) or the "Reproducibility Project" (Collaboration, 2012). In fact, the data also indicate that these projects may consider starting with replicating findings published in high-ranking journals.

Given all of the above evidence, it is therefore not surprising that journal rank is also a strong predictor of the rate of retractions (**Figure 1D**) (Liu, 2006; Cokol et al., 2007; Fang and Casadevall, 2011).

SOCIAL PRESSURE AND JOURNAL RANK

There are thus several converging lines of evidence which indicate that publications in high ranking journals are not only more likely to be fraudulent than articles in lower ranking journals, but also more likely to present discoveries which are less reliable (i.e., are inflated, or cannot subsequently be replicated). Some of the sociological mechanisms behind these correlations have been documented, such as pressure to publish (preferably positive results in high-ranking journals), leading to the potential for decreased ethical standards (Anderson et al., 2007) and increased publication bias in highly competitive fields (Fanelli, 2010). The general increase in competitiveness, and the precariousness of scientific careers (Shapin, 2008), may also lead to an increased publication bias across the sciences (Fanelli, 2011). This evidence supports earlier propositions about social pressure being a major factor driving misconduct and publication bias (Giles, 2007), eventually culminating in retractions in the most extreme cases.
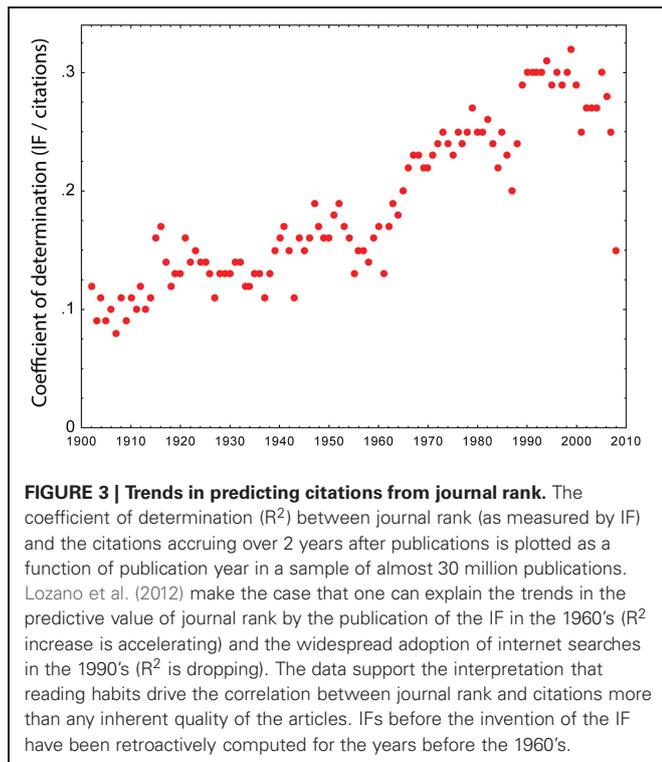
That being said, it is clear that the correlation between journal rank and retraction rate is likely too strong (coefficient of determination of 0.77; data from Fang and Casadevall, 2011) to be explained exclusively by the decreased reliability of the research published in high ranking journals. Probably, additional factors contribute to this effect. For instance, one such factor may be the greater visibility of publications in these journals, which is both one of the incentives driving publication bias, and a likely underlying cause for the detection of error or misconduct with the eventual retraction of the publications as a result (Cokol et al., 2007). Conversely, the scientific community may also be less concerned about incorrect findings published in more obscure journals. With respect to the latter, the finding that the large majority of retractions come from the numerous lower-ranking journals (Fang et al., 2012) reveals that publications in lower ranking journals are scrutinized and, if warranted, retracted. Thus, differences in scrutiny are likely to be only a contributing factor and not an exclusive explanation, either. With respect to the former, visibility effects in general can be quantified by measuring citation rates between journals, testing the assumption that if higher visibility were a contributing factor to retractions, it must also contribute to citations.

JOURNAL RANK AND STUDY IMPACT

Thus far we have presented evidence that research published in high-ranking journals may be less reliable compared with publications in lower-ranking journals. Nevertheless, there is a strong common perception that high-ranking journals publish "better" or "more important" science, and that the IF captures this well (Gordon, 1982; Saha et al., 2003). The assumption is that high-ranking journals are able to be highly selective and publish only the most important, novel and best-supported scientific discoveries, which will then, as a consequence of their quality, go on to be highly cited (Young et al., 2008). One way to reconcile this common perception with the data would be that, while journal rank may be indicative of a minority of unreliable publications, it may also (or more strongly) be indicative of the importance of the majority of remaining, reliable publications. Indeed, a recent study on clinical trial meta-analyses found that a measure for the novelty of a clinical trial's main outcome did correlate significantly with journal rank (Evangelou et al., 2012). Compared to this relatively weak correlation (with all coefficients of determination lower than 0.1), a stronger correlation was reported for journal rank and expert ratings of importance (Allen et al., 2009). In this study, the journal in which the study had appeared was not masked, thus not excluding the strong correlation between subjective journal rank and journal quality as a confounding factor. Nevertheless, there is converging evidence from two studies that journal rank is indeed indicative of a publication's perceived importance.

Beyond the importance or novelty of the research, there are three additional reasons why publications in high-ranking journals might receive a high number of citations. First, publications in high-ranking journals achieve greater exposure by virtue not only of the larger circulation of the journal in which they appear, but also of the more prominent media attention (Gonon et al., 2012). Second, citing high-ranking publications in one's own publication may increase its perceived value. Third, the novel, surprising, counter-intuitive or controversial findings often published in high-ranking journals, draw citations not only from follow-up studies but also from news-type articles in scholarly journals reporting and discussing the discovery. Despite these four factors, which would suggest considerable effects of journal rank on future citations, it has been established for some time that the actual effect of journal rank is measurable, but nowhere near as substantial as indicated (Seglen, 1994, 1997; Callaham, 2002; Chow et al., 2007; Kravitz and Baker, 2011; Hegarty and Walton, 2012; Finardi, 2013) and as one would expect if visibility were the exclusive factor driving retractions. In fact, the average effect sizes roughly approach those for journal rank and unreliability, cited above.

The data presented in a recent analysis of the development of these correlations between IF-based journal rank and future citations over the period from 1902–2009 (with IFs before the 1960's computed retroactively) reveal two very informative trends (**Figure 3**, data from (Lozano et al., 2012). First, while the predictive power of journal rank remained very low for the entire first two thirds of the twentieth century, it started to slowly increase shortly after the publication of the first IF data in the 1960's. This correlation kept increasing until the second interesting trend

FIGURE 3 | Trends in predicting citations from journal rank. The coefficient of determination (R^2) between journal rank (as measured by IF) and the citations accruing over 2 years after publications is plotted as a function of publication year in a sample of almost 30 million publications. Lozano et al. (2012) make the case that one can explain the trends in the predictive value of journal rank by the publication of the IF in the 1960's (R^2 increase is accelerating) and the widespread adoption of internet searches in the 1990's (R^2 is dropping). The data support the interpretation that reading habits drive the correlation between journal rank and citations more than any inherent quality of the articles. IFs before the invention of the IF have been retroactively computed for the years before the 1960's.

emerged with the advent of the internet and keyword-search engines in the 1990's, from which time on it fell back to pre-1960's levels until the end of the study period in 2009. Overall, consistent with the citation data already available, the coefficient of determination between journal rank and citations was always in the range of ∼0.1 to 0.3 (i.e., quite low). It thus appears that indeed a small but significant correlation between journal rank and future citations can be observed. Moreover, the data suggest that most of this small effect stems from visibility effects due to the influence of the IF on reading habits (Lozano et al., 2012), rather than from factors intrinsic to the published articles (see data cited above). However, the correlation is so weak that it cannot alone account for the strong correlation between retractions and journal rank, but instead requires additional factors, such as the increased unreliability of publications in high ranking journals cited above. Supporting these weak correlations between journal rank and future citations are data reporting classification errors (i.e., whether a publication received too many or too few citations with regard to the rank of the journal it was published in) at or exceeding 30% (Starbuck, 2005; Chow et al., 2007; Singh et al., 2007; Kravitz and Baker, 2011). In fact, these classification errors, in conjunction with the weak citation advantage, render journal rank practically useless as an evaluation signal, even if there was no indication of less reliable science being published in high ranking journals.

The only measure of citation count that does correlate strongly with journal rank (negatively) is the number of articles without any citations at all (Weale et al., 2004), supporting the argument that fewer articles in high-ranking journals go unread. Thus, there is quite extensive evidence arguing for the strong correlation

between journal rank and retraction rate to be mainly due to two factors: there is direct evidence that the social pressures to publish in high ranking journals increases the unreliability, intentional or not, of the research published there. There is more indirect evidence, derived mainly from citation data, indicating that increased visibility of publications in high ranking journals may potentially contribute to increased error-detection in these journals. With several independent measures failing to provide compelling evidence that journal rank is a reliable predictor of scientific impact or quality, and other measures indicating that journal rank is at least equally if not more predictive of low reliability, the central role of journal rank in modern science deserves close scrutiny.

PRACTICAL CONSEQUENCES OF JOURNAL RANK

Even if a particular study has been performed to the highest standards, the quest for publication in high-ranking journals slows down the dissemination of science and increases the burden on reviewers, by iterations of submissions and rejections cascading down the hierarchy of journal rank (Statzner and Resh, 2010; Kravitz and Baker, 2011; Nosek and Bar-Anan, 2012). A recent study seems to suggest that such rejections eventually improve manuscripts enough to yield measurable citation benefits (Calcagno et al., 2012). However, the effect size of such resubmissions appears to be of the order of 0.1 citations per article, a statistically significant but, in practical terms, negligible effect. This conclusion is corroborated by an earlier study which failed to find any such effect (Nosek and Bar-Anan, 2012). Moreover, with peer-review costs estimated in excess of 2.2 billion € (US$ ∼2.8b) annually (Research Information Network, 2008), the resubmission cascade contributes to the already rising costs of journal rank: the focus on journal rank has allowed corporate publishers to keep their most prestigious journals closed-access and to increase subscription prices (Kyrillidou et al., 2012), creating additional barriers to the dissemination of science. The argument from highly selective journals is that their per-article cost would be too high for author processing fees, which may be up to 37,000€ (US$ 48,000) for the journal Nature (House of Commons, 2004). There is also evidence from one study in economics suggesting that journal rank can contribute to suppression of interdisciplinary research (Rafols et al., 2012), keeping disciplines separate and isolated.

Finally, the attention given to publication in high-ranking journals may distort the communication of scientific progress, both inside and outside of the scientific community. For instance, the recent discovery of a "Default-Mode Network" in rodent brains was, presumably, made independently by two different sets of neuroscientists and published only a few months apart (Upadhyay et al., 2011; Lu et al., 2012). The later, but not the earlier, publication (Lu et al., 2012) was cited in a subsequent high-ranking publication (Welberg, 2012). Despite both studies largely reporting identical findings (albeit, perhaps, with different quality), the later report has garnered 19 citations, while the earlier one only 5, at the time of this writing. We do not know of any empirical studies quantitatively addressing this particular effect of journal rank. However, a similar distortion due to

selective attention to publications in high-ranking journals has been reported in a study on medical research. This study found media reporting to be distorted, such that once initial findings in higher-ranking journals have been refuted by publications in lower ranking journals (a case of decline effect), they do not receive adequate media coverage (Gonon et al., 2012).

IMPACT FACTOR—NEGOTIATED, IRREPRODUCIBLE, AND UNSOUND

The IF is a metric for the number of citations to articles in a journal (the numerator), normalized by the number of articles in that journal (the denominator). However, there is evidence that IF is, at least in some cases, not calculated but negotiated, that it is not reproducible, and that, even if it were reproducibly computed, the way it is derived is not mathematically sound. The fact that publishers have the option to negotiate how their IF is calculated is well-established—in the case of PLoS Medicine, the negotiation range was between 2 and about 11 (Editorial, 2006). What is negotiated is the denominator in the IF equation (i.e., which published articles which are counted), given that all citations count toward the numerator whether they result from publications included in the denominator or not. It has thus been public knowledge for quite some time now that removing editorials and News-and-Views articles from the denominator (so called "front-matter") can dramatically alter the resulting IF (Moed and Van Leeuwen, 1995, 1996; Baylis et al., 1999; Garfield, 1999; Adam, 2002; Editorial, 2005; Hernán, 2009). While these IF negotiations are rarely made public, the number of citations (numerator) and published articles (denominator) used to calculate IF are accessible via Journal Citation Reports. This database can be searched for evidence that the IF has been negotiated. For instance, the numerator and denominator values for Current Biology in 2002 and 2003 indicate that while the number of citations remained relatively constant, the number of published articles dropped. This decrease occurred after the journal was purchased by Cell Press (an imprint of Elsevier), despite there being no change in the layout of the journal. Critically, the arrival of a new publisher corresponded with a retrospective change in the denominator

Table 1 | Thomson Reuters' IF calculations for the journal "Current Biology" in the years 2002/2003.

Journal: current biology	Published items 2000	Published items 2001	Published items 2002	Sum published items	Citations in preceding 2 years	IF
JCR science edition 2002	504	528	n.c.	1032	7231	7.007
JCR science edition 2003	n.c.	300	334	634	7551	11.910

Most of the rise in IF is due to the reduction in published items.

Note the discrepancy between the number of items published in 2001 between the two consecutive JCR Science Editions.

*n.c.: year not covered by this edition. Raw data see **Figure A1**.*

used to calculate IF (**Table 1**). Similar procedures raised the IF of FASEB Journal from 0.24 in 1988 to 18.3 in 1989, when conference abstracts ceased to count toward the denominator (Baylis et al., 1999).

In an attempt to test the accuracy of the ranking of some of their journals by IF, Rockefeller University Press purchased access to the citation data of their journals and some competitors. They found numerous discrepancies between the data they received and the published rankings, sometimes leading to differences of up to 19% (Rossner et al., 2007). When asked to explain this discrepancy, Thomson Reuters replied that they routinely use several different databases and had accidentally sent Rockefeller University Press the wrong one. Despite this, a second database sent also did not match the published records. This is only one of a number reported errors and inconsistencies (Reedijk, 1998; Moed et al., 1996).

It is well-known that citation data are strongly left-skewed, meaning that a small number of publications receive a large number of citations, while most publications receive very few (Seglen, 1992, 1997; Weale et al., 2004; Editorial, 2005; Chow et al., 2007; Rossner et al., 2007; Taylor et al., 2008; Kravitz and Baker, 2011). The use of an arithmetic mean as a measure of central tendency on such data (rather than, say, the median) is clearly inappropriate, but this is exactly what is used in the IF calculation. The International Mathematical Union reached the same conclusion in an analysis of the IF (Adler et al., 2008). A recent study correlated the median citation frequency in a sample of 100 journals with their 2-year IF and found a very strong correlation, which is expected due to the similarly left-skewed distributions in most journals (Editorial, 2013). However, at the time of this writing, it is not known if using the median (instead of the mean) improves any of the predominantly weak predictive properties of journal rank. Complementing the specific flaws just mentioned, a recent, comprehensive review of the bibliometric literature lists various additional shortcomings of the IF more generally (Vanclay, 2011).

CONCLUSIONS

While at this point it seems impossible to quantify the relative contributions of the different factors influencing the reliability of scientific publications, the current empirical literature on the effects of journal rank provides evidence supporting the following four conclusions: (1) journal rank is a weak to moderate predictor of utility and perceived importance; (2) journal rank is a moderate to strong predictor of both intentional and unintentional scientific unreliability; (3) journal rank is expensive, delays science and frustrates researchers; and, (4) journal rank as established by IF violates even the most basic scientific standards, but predicts subjective judgments of journal quality.

CAVEATS

While our latter two conclusions appear uncontroversial, the former two are counter-intuitive and require explanation. Weak correlations between future citations and journal rank based on IF may be caused by the poor statistical properties of the IF. This explanation could (and should) be tested by using any of the existing alternative ranking tools available (such as Thomson

Reuters' Eigenfactor, Scopus' SCImagoJournalRank, or Google's Scholar Metrics etc.) and computing correlations with the metrics discussed above. However, a recent analysis shows a high correlation between these ranks, so no large differences would be expected (Lopez-Cozar and Cabezas-Clavijo, 2013). Alternatively, one can choose other important metrics and compute which journals score particularly high on these. Either way, since the IF reflects the common perception of journal hierarchies rather well (Gordon, 1982; Saha et al., 2003; Yue et al., 2007; Sønderstrup-Andersen and Sønderstrup-Andersen, 2008), any alternative hierarchy that would better reflect article citation frequencies might violate this intuitive sense of journal rank, as different ways to compute journal rank lead to different hierarchies (Wagner, 2011). Both alternatives thus challenge our subjective journal ranking. To put it more bluntly, if perceived importance and utility were to be discounted as indirect proxies of quality, while retraction rate, replicability, effect size overestimation, correct sample sizes, crystallographic quality, sound methodology and so on counted as more direct measures of quality, then inversing the current IF-based journal hierarchy would improve the alignment of journal rank for most and have no effect on the rest of these more direct measures of quality.

The subjective journal hierarchy also leads to a circularity that confounds many empirical studies. That is, authors use journal rank, in part, to make decisions of where to submit their manuscripts, such that well-performed studies yielding groundbreaking discoveries with general implications are preferentially submitted to high-ranking journals. Readers, in turn, expect only to read about such articles in high-ranking journals, leading to the exposure and visibility confounds discussed above and at length in the cited literature. Moreover, citation practices and methodological standards vary in different scientific fields, potentially distorting both the citation and reliability data. Given these confounds one might expect highly varying and often inconclusive results. Despite this, the literature contains evidence for associations between journal rank and measures of scientific impact (e.g., citations, importance, and unread articles), but also contains at least equally strong, consistent effects of journal rank predicting scientific unreliability (e.g., retractions, effect size, sample size, replicability, fraud/misconduct, and methodology). Neither group of studies can thus be easily dismissed, suggesting that the incentives journal rank creates for the scientific community (to submit either their best or their most unreliable work to the most high-ranking journals) at best cancel each other out. Such unintended consequences are well-known from other fields where metrics are applied (Hauser and Katz, 1998).

Therefore, while there are concerns not only about the validity of the IF as the metric of choice for establishing journal rank but also about confounding factors complicating the interpretation of some of the data, we find, in the absence of additional data, that these concerns do not suffice to substantially question our conclusions, but do emphasize the need for future research.

POTENTIAL LONG-TERM CONSEQUENCES OF JOURNAL RANK

Taken together, the reviewed literature suggests that using journal rank is unhelpful at best and unscientific at worst. In our view, IF generates an illusion of exclusivity and prestige based on an assumption that it predicts scientific quality, which is not supported by empirical data. As the IF aligns well with intuitive notions of journal hierarchies (Gordon, 1982; Saha et al., 2003; Yue et al., 2007), it receives insufficient scrutiny (Frank, 2003) (perhaps a case of confirmation bias). The one field in which journal rank is scrutinized is bibliometrics. We have reviewed the pertinent empirical literature to supplement the largely argumentative discussion on the opinion pages of many learned journals (Moed and Van Leeuwen, 1996; Lawrence, 2002, 2007, 2008; Bauer, 2004; Editorial, 2005; Giles, 2007; Taylor et al., 2008; Todd and Ladle, 2008; Tsikliras, 2008; Adler and Harzing, 2009; Garwood, 2011; Schooler, 2011; Brumback, 2012; Sarewitz, 2012) with empirical data. Much like dowsing, homeopathy or astrology, journal rank seems to appeal to subjective impressions of certain effects, but these effects disappear as soon as they are subjected to scientific scrutiny.

In our understanding of the data, the social and psychological influences described above are, at least to some extent, generated by journal rank itself, which in turn may contribute to the observed decline effect and rise in retraction rate. That is, systemic pressures on the author, rather than increased scrutiny on the part of the reader, inflate the unreliability of much scientific research. Without reform of our publication system, the incentives associated with increased pressure to publish in high-ranking journals will continue to encourage scientists to be less cautious in their conclusions (or worse), in an attempt to market their research to the top journals (Anderson et al., 2007; Giles, 2007; Shapin, 2008; Munafò et al., 2009; Fanelli, 2010). This is reflected in the decline in null results reported across disciplines and countries (Fanelli, 2011), and corroborated by the findings that much of the increase in retractions may be due to misconduct (Steen, 2011b; Fang et al., 2012), and that much of this misconduct occurs in studies published high-ranking journals (Steen, 2011a; Fang et al., 2012). Inasmuch as journal rank guides the appointment and promotion policies of research institutions, the increasing rate of misconduct that has recently been observed may prove to be but the beginning of a pandemic: It is conceivable that, for the last few decades, research institutions world-wide may have been hiring and promoting scientists who excel at marketing their work to top journals, but who are not necessarily equally good at conducting their research. Conversely, these institutions may have purged excellent scientists from their ranks, whose marketing skills did not meet institutional requirements. If this interpretation of the data is correct, a generation of excellent marketers (possibly, but not necessarily, also excellent scientists) now serve as the leading figures and role models of the scientific enterprise, constituting another potentially major contributing factor to the rise in retractions.

The implications of the data presented here go beyond the reliability of scientific publications—public trust in science and scientists has been in decline for some time in many countries (Nowotny, 2005; European Commission, 2010; Gauchat, 2010), dramatically so in some sections of society (Gauchat, 2012), culminating in the sentiment that scientists are nothing more than yet another special interest group (Miller, 2012; Sarewitz, 2013). In the words of Daniel Sarewitz: "Nothing will

corrode public trust more than a creeping awareness that scientists are unable to live up to the standards that they have set for themselves" (Sarewitz, 2012). The data presented here prompt the suspicion that the corrosion has already begun and that journal rank may have played a part in this decline as well.

ALTERNATIVES

Alternatives to journal rank exist—we now have technology at our disposal which allows us to perform all of the functions journal rank is currently supposed to perform in an unbiased, dynamic way on a per-article basis, allowing the research community greater control over selection, filtering, and ranking of scientific information (Hönekopp and Khan, 2011; Kravitz and Baker, 2011; Lin, 2012; Priem et al., 2012; Roemer and Borchardt, 2012; Priem, 2013). Since there is no technological reason to continue using journal rank, one implication of the data reviewed here is that we can instead use current technology and remove the need for a journal hierarchy completely. As we have argued, it is not only technically obsolete, but also counter-productive and a potential threat to the scientific endeavor. We therefore would favor bringing scholarly communication back to the research institutions in an archival publication system in which both software, raw data and their text descriptions are archived and made accessible, after peer-review and with scientifically-tested metrics accruing reputation in a constantly improving reputation system (Eve, 2012). This reputation system would be subjected to the same standards of scientific scrutiny as are commonly applied to all scientific matters and evolve to minimize gaming and maximize the alignment of researchers' interests with those of science [which are currently misaligned (Nosek et al., 2012)]. Only an elaborate ecosystem of a multitude of metrics can provide the flexibility to capitalize on the small fraction of the multi-faceted scientific output that is actually quantifiable. Such an ecosystem would evolve such that the only evolutionary stable strategy is to try and do the best science one can.

The currently balkanized literature, with a lack of interoperability and standards as one of its many detrimental, unintended consequences, prevents the kind of innovation that gave rise to the discover functions of Amazon or eBay, the social networking functions of Facebook or Reddit and course the sort and search functions of Google—all technologies virtually every scientist uses regularly for all activities but science. Thus, fragmentation and the resulting lack of access and interoperability are among the main underlying reasons why journal rank has not yet been replaced by more scientific evaluation options, despite widespread access to article-level metrics today. With an openly accessible scholarly literature standardized for interoperability, it would of course still be possible to pay professional editors to select publications, as is the case now, but after publication. These editors would then actually compete with each other for paying customers, accumulating track records for selecting (or missing) the most important discoveries. Likewise, virtually any functionality the current system offers would easily be replicable in the system we envisage. However, above and beyond replicating current functionality, an open, standardized scholarly literature would place any and all thinkable scientific metrics only a few lines of code away, offering the possibility of a truly open evaluation system where any hypothesis can be tested. Metrics, social networks and intelligent software then can provide each individual user with regular, customized updates on the most relevant research. These updates respond to the behavior of the user and learn from and evolve with their preferences. With openly accessible, interoperable literature, data and software, agents can be developed that independently search for hypotheses in the vast knowledge accumulating there. But perhaps most importantly, with an openly accessible database of science, innovation can thrive, bringing us features and ideas nobody can think of today and nobody will ever be capable of imagining, if we do not bring the products of our labor back under our own control. It was the hypertext transfer protocol (http) standard that spurred innovation and made the internet what it is today. What is required is the equivalent of http for scholarly literature, data and software.

Funds currently spent on journal subscriptions could easily suffice to finance the initial conversion of scholarly communication, even if only as long-term savings. One avenue to move in this direction may be the recently announced Episcience Project (Van Noorden, 2013). Other solutions certainly exist (Bachmann, 2011; Birukou et al., 2011; Kravitz and Baker, 2011; Kreiman and Maunsell, 2011; Zimmermann et al., 2011; Beverungen et al., 2012; Florian, 2012; Ghosh et al., 2012; Hartshorne and Schachner, 2012; Hunter, 2012; Ietto-Gillies, 2012; Kriegeskorte, 2012; Kriegeskorte et al., 2012; Lee, 2012; Nosek and Bar-Anan, 2012; Pöschl, 2012; Priem and Hemminger, 2012; Sandewall, 2012; Walther and Van den Bosch, 2012; Wicherts et al., 2012; Yarkoni, 2012), but the need for an alternative system is clearly pressing (Casadevall and Fang, 2012). Given the data we surveyed above, almost anything appears superior to the status quo.

REFERENCES

Adam, D. (2002). The counting house. *Nature* 415, 726–729. doi: 10.1038/415726a

Adler, N. J., and Harzing, A.-W. (2009). When knowledge wins: transcending the sense and nonsense of academic rankings. *Acad. Manag. Learn. Edu.* 8, 72–95. doi: 10.5465/AMLE.2009.37012181

Adler, R., Ewing, J., and Taylor, P. (2008). *Joint Committee on Quantitative Assessment of Research: Citation Statistics (A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathemat.* Available online at: http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf

Allen, L., Jones, C., Dolby, K., Lynn, D., and Walport, M. (2009). Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS ONE* 4:8. doi: 10.1371/journal.pone.0005910

Anderson, M. S., Martinson, B. C., and De Vries, R. (2007). Normative dissonance in science: results from a national survey of u.s. Scientists. *JERHRE* 2, 3–14. doi: 10.1525/jer.2007.2.4.3

Bachmann, T. (2011). Fair and open evaluation may call for temporarily hidden authorship, caution when counting the votes, and transparency of the full pre-publication procedure. *Front. Comput. Neurosci.* 5:61. doi: 10.3389/fncom.2011.00061

Bain, C. R., and Myles, P. S. (2005). Relationship between journal impact factor and levels of evidence in anaesthesia. *Anaesth. Intensive Care* 33, 567–570.

Baker, M. (2012). Independent Labs to Verify High-Profile Papers. *Nature News.* Available online at: http://www.nature.com/doifinder/10.1038/nature.2012.11176 doi: 10.1038/nature.2012.11176

Bauer, H. H. (2004). Science in the 21st century?: knowledge monopolies and research cartels. *J. Scient. Explor.* 18, 643–660.

Baylis, M., Gravenor, M., and Kao, R. (1999). Sprucing up one's impact factor. *Nature* 401, 322.

Begley, C. G., and Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533. doi: 10.1038/483531a

Bertamini, M., and Munafo, M. R. (2012). Bite-size science and its undesired side effects. *Perspect. Psychol. Sci.* 7, 67–71. doi: 10.1177/1745691611429353

Beverungen, A., Bohm, S., and Land, C. (2012). The poverty of journal publishing. *Organization* 19, 929–938. doi: 10.1177/1350508412448858

Birukou, A., Wakeling, J. R., Bartolini, C., Casati, F., Marchese, M., Mirylenka, K., et al. (2011). Alternatives to peer review: novel approaches for research evaluation. *Front. Comput. Neurosci.* 5:56. doi: 10.3389/fncom.2011.00056

Brown, E. N., and Ramaswamy, S. (2007). Quality of protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 63, 941–950. doi: 10.1107/S0907444907033847

Brumback, R. A. (2012). "3.. 2.. 1.. Impact [factor]: target [academic career] destroyed!": just another statistical casualty. *J. Child Neurol.* 27, 1565–1576. doi: 10.1177/0883073812465014

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Calcagno, V., Demoinet, E., Gollner, K., Guidi, L., Ruths, D., and De Mazancourt, C. (2012). Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science (N.Y.)* 338, 1065–1069. doi: 10.1126/science.1227833

Callaham, M. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 287, 2847–2850. doi: 10.1001/jama.287.21.2847

Carpenter, S. (2012). Psychology's bold initiative. *Science* 335, 1558–1561. doi: 10.1126/science.335.6076.1558

Casadevall, A., and Fang, F. C. (2012). Reforming science: methodological and cultural reforms. *Inf. Immun.* 80, 891–896. doi: 10.1128/IAI.06183-11

Chow, C. W., Haddad, K., Singh, G., and Wu, A. (2007). On using journal rank to proxy for an article's contribution or value. *Issues Account. Edu.* 22, 411–427. doi: 10.2308/iace.2007.22.3.411

Cokol, M., Iossifov, I., Rodriguez-Esteban, R., and Rzhetsky, A. (2007). How many scientific papers should be retracted? *EMBO Reports* 8, 422–423. doi: 10.1038/sj.embor.7400970

Collaboration, O. S. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspect. Psychol. Sci.* 7, 657–660. doi: 10.1177/1745691612462588

Curry, S. (2009). Eye-opening Access. *Occam's Typewriter: Reciprocal Space.* Available online at: http://occamstypewriter.org/scurry/2009/03/27/eye_opening_access/

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., et al. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 3:e3081. doi: 10.1371/journal.pone.0003081

Editorial. (2005). Not-so-deep impact. *Nature* 435, 1003–1004. doi: 10.1038/4351003b

Editorial. (2006). The impact factor game. *It is time to find a better way to assess the scientific literature.* PLoS Med. 3:e291. doi: 10.1371/journal.pmed.0030291

Editorial. (2012). The well-behaved scientist. *Science* 335, 285–285.

Editorial. (2013). Beware the impact factor. *Nat. Materials* 12:89. doi: 10.1038/nmat3566

European Commission. (2010). Science and Technology Report. 340 / Wave 73.1 – TNS Opinion and Social.

Evangelou, E., Siontis, K. C., Pfeiffer, T., and Ioannidis, J. P. A. (2012). Perceived information gain from randomized trials correlates with publication in high-impact factor journals. *J. Clin. Epidemiol.* 65, 1274–1281. doi: 10.1016/j.jclinepi.2012.06.009

Eve, M. P. (2012). Tear it down, build it up: the research output team, or the library-as-publisher. *Insights UKSG J.* 25, 158–162. doi: 10.1629/2048-7754.25.2.158

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4:e5738. doi: 10.1371/journal.pone.0005738

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS ONE* 5:e10271. doi: 10.1371/journal.pone.0010271

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics* 90, 891–904. doi: 10.1007/s11192-011-0494-7

Fang, F. C., and Casadevall, A. (2011). Retracted science and the retraction index. *Inf. Immun.* 79, 3855–3859. doi: 10.1128/IAI.05661-11

Fang, F. C., Steen, R. G., and Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17028–17033. doi: 10.1073/pnas.1212247109

Finardi, U. (2013). Correlation between journal impact factor and citation performance: an experimental study. *J. Inf.* 7, 357–370. doi: 10.1016/j.joi.2012.12.004

Florian, R. V. (2012). Aggregating post-publication peer reviews and ratings. *Front. Comput. Neurosci.* 6:31. doi: 10.3389/fncom.2012.00031

Frank, M. (2003). Impact factors: arbiter of excellence? *JMLA* 91, 4–6.

Garfield, E. (1999). Journal impact factor: a brief review. *CMAJ* 161, 979–980.

Garwood, J. (2011). A conversation with Peter Lawrence, Cambridge. "The Heart of Research is Sick". *LabTimes* 2, 24–31.

Gauchat, G. (2010). The cultural authority of science: public trust and acceptance of organized science. *Pub. Understand. Sci.* 20, 751–770. doi: 10.1177/0963662510365246

Gauchat, G. (2012). Politicization of Science in the public sphere: a study of public trust in the United States, 1974 to 2010. *Am. Sociol. Rev.* 77, 167–187. doi: 10.1177/0003122412438225

Ghosh, S. S., Klein, A., Avants, B., and Millman, K. J. (2012). Learning from open source software projects to improve scientific review. *Front. Comput. Neurosci.* 6:18. doi: 10.3389/fncom.2012.00018

Giles, J. (2007). Breeding cheats. *Nature* 445, 242–243. doi: 10.1038/445242a

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards Grease the way through the publication bottleneck but undermine science. *Perspect. Psychol. Sci.* 7, 562–571. doi: 10.1177/1745691612457576

Goldacre, B. (2011). I Foresee that Nobody Will Do Anything About this Problem. *Bad Science.* Available online at: http://www.badscience.net/2011/04/i-foresee-that-nobody-will-do-anything-about-this-problem

Gonon, F., Konsman, J. -P., Cohen, D., and Boraud, T. (2012). Why most biomedical findings Echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. *PLoS ONE* 7:e44275. doi: 10.1371/journal.pone.0044275

Gordon, M. D. (1982). Citation ranking versus subjective evaluation in the determination of journal hierachies in the social sciences. *J. Am. Soc. Inf. Sci.* 33, 55–57. doi: 10.1002/asi.4630330109

Hamilton, J. (2011). Debunked Science: Studies Take Heat In 2011. *NPR.* Available online at: http://www.npr.org/2011/12/29/144431640/debunked-science-studies-take-heat-in-2011

Hartshorne, J. K., and Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Front. Comput. Neurosci.* 6:8. doi: 10.3389/fncom.2012.00008

Hauser, J. R., and Katz, G. M. (1998). Metrics: you are what you measure! *Eur. Manag. J.* 16, 517–528. doi: 10.1016/S0263-2373(98)00029-2

Hegarty, P., and Walton, Z. (2012). The consequences of predicting scientific impact in psychology using journal impact factors. *Perspect. Psychol. Sci.* 7, 72–78. doi: 10.1177/1745691611429356

Hernán, M. A. (2009). Impact factor: a call to reason. *Epidemiology* 20, 317–318; discussion 319–20. doi: 10.1097/EDE.0b013e31819ed4a6

Hönekopp, J., and Khan, J. (2011). Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics* 90, 843–853. doi: 10.1007/s11192-011-0551-2

House of Commons. (2004). Scientific publications: free for all? *Tenth Report of Session 2003-2004*, vol II: Written evidence, Appendix 138.

Hunter, J. (2012). Post-publication peer review: opening up scientific conversation. *Front. Comput. Neurosci.* 6:63. doi: 10.3389/fncom.2012.00063

Ietto-Gillies, G. (2012). The evaluation of research papers in the XXI century. *The open peer discussion system of the world economics association. Front. Comput. Neurosci.* 6:54. doi: 10.3389/fncom.2012.00054

Ioannidis, J. P. A. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294, 218–228. doi: 10.1001/jama.294.2.218

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

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspect. Psychol. Sci.* 7, 645–654. doi: 10.1177/1745691612464056

Ioannidis, J. P. A., and Panagiotou, O. A. (2011). Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* 305, 2200–2210. doi: 10.1001/jama.2011.713

Kelly, C. D. (2006). Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. *Q. Rev. Biol.* 81, 221–236. doi: 10.1086/506236

Kravitz, D. J., and Baker, C. I. (2011). Toward a new model of scientific publishing: discussion and a proposal. *Front.*

Comput. Neurosci. 5:55. doi: 10.3389/fncom.2011.00055

Kreiman, G., and Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Front. Comput. Neurosci.* 5:48. doi: 10.3389/fncom.2011.00048

Kriegeskorte, N. (2012). Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. *Front. Comput. Neurosci.* 6:79. doi: 10.3389/fncom.2012.00079

Kriegeskorte, N., Walther, A., and Deca, D. (2012). An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. *Front. Comput. Neurosci.* 6:94. doi: 10.3389/fncom.2012.00094

Kyrillidou, M., Morris, S., and Roebuck, G. (2012). ARL Statistics. *American Research Libraries Digital Publications.* Available online at: http://publications.arl.org/ARL_Statistics

Lau, S. L., and Samman, N. (2007). Levels of evidence and journal impact factor in oral and maxillofacial surgery. *Int. J. Oral Maxillofac. Surg.* 36, 1–5. doi: 10.1016/j.ijom.2006.10.008

Lawrence, P. (2008). Lost in publication: how measurement harms science. *Ethics Sci. Environ. Politics* 8, 9–11. doi: 10.3354/esep00079

Lawrence, P. A. (2002). Rank injustice. *Nature* 415, 835–836. doi: 10.1038/415835a

Lawrence, P. A. (2007). The mismeasurement of science. *Curr. Biol.* 17, R583–R585. doi: 10.1016/j.cub.2007.06.014

Lee, C. (2012). Open peer review by a selected-papers network. *Front. Comput. Neurosci.* 6:1. doi: 10.3389/fncom.2012.00001

Lehrer, J. (2010). *The Decline Effect and the Scientific Method.* New Yorker. Available online at: http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer

Lin, J. (2012). Cracking Open the Scientific Process: "Open Science" Challenges *Journal Tradition With Web Collaboration.* New York Times. Available online at: http://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html?pagewanted=all

Liu, S. (2006). Top journals' top retraction rates. *Sci. Ethics* 1, 92–93.

Lopez-Cozar, E. D., and Cabezas-Clavijo, A. (2013). Ranking journals: could Google scholar metrics be an alternative to journal citation reports and *Scimago journal rank?* ArXiv 1303.5870, 26. doi: 10.1087/20130206

Lozano, G. A., Larivière, V., and Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *J. Am. Soc. Inf. Sci. Technol.* 63, 2140–2145. doi: 10.1002/asi.22731

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., and Yang, Y. (2012). Rat brains also have a default mode network. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3979–3984. doi: 10.1073/pnas.1200506109

Makel, M. C., Plucker, J. A., and Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspect. Psychol. Sci.* 7, 537–542. doi: 10.1177/1745691612460688

Miller, K. R. (2012). America's Darwin Problem. *Huffington Post.* Available online at: http://www.huffingtonpost.com/kenneth-r-miller/darwin-day-evolution_b_1269191.html

Moed, H. F., and Van Leeuwen, T. N. (1995). Improving the accuracy of institute for scientific information's journal impact factors. *J. Am. Soc. Inf. Sci.* 46, 461–467. doi: 10.1002/(SICI)1097-4571(199507)46:6<461::AID-ASI5>3.0.CO;2-G

Moed, H. F., and Van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature* 381, 186. doi: 10.1038/381186a0

Moed, H. F., Van Leeuwen, T. N., and Reedijk, J. (1996). A critical analysis of the journal impact factors of Angewandte Chemie and the journal of the American Chemical Society inaccuracies in published impact factors based on overall citations only. *Scientometrics* 37, 105–116. doi: 10.1007/BF02093487

Møller, A. P., and Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends Ecol. Evol.* 16, 580–586. doi: 10.1016/S0169-5347(01)02235-2

Møller, A. P., Thornhill, R., and Gangestad, S. W. (2005). Direct and indirect tests for publication bias: asymmetry and sexual selection. *Anim. Behav.* 70, 497–506. doi: 10.1016/j.anbehav.2004.11.017

Munafò, M. R., Freimer, N. B., Ng, W., Ophoff, R., Veijola, J., Miettunen, J., et al. (2009). 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. *Am. J. Med. Genet. B Neuropsychiat. Genet.* 150B, 271–281. doi: 10.1002/ajmg.b.30808

Munafò, M. R., Matheson, I. J., and Flint, J. (2007). Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence

of publication bias. *Mol. Psychiatry* 12, 454–461. doi: 10.1038/sj.mp.4001938

Nath, S. B., Marcus, S. C., and Druss, B. G. (2006). Retractions in the research literature: misconduct or mistakes? *Med. J. Aust.* 185, 152–154.

Nosek, B. A., and Bar-Anan, Y. (2012). Scientific Utopia: I. *opening scientific communication. Psychol. Inquiry* 23, 217–243. doi: 10.1080/1047840X.2012.692215

Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific Utopia: II. *restructuring incentives and practices to promote truth over publishability. Perspect. Psychol. Sci.* 7, 615–631. doi: 10.1177/1745691612459058

Nowotny, H. (2005). Science and society. *High- and low-cost realities for science and society. Science (N.Y.)* 308, 1117–1118. doi: 10.1126/science.1113825

Obremskey, W. T., Pappas, N., Attallah-Wasif, E., Tornetta, P., and Bhandari, M. (2005). Level of evidence in orthopaedic journals. *J. Bone Joint Surg. Am.* 87, 2632–2638. doi: 10.2106/JBJS.E.00370

Palmer, A. R. (2000). Quasi-replication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Ann. Rev. Ecol. Syst.* 31, 441–480. doi: 10.1146/annurev.ecolsys.31.1.441

Popper, K. (1995). In Search of A Better World: Lectures and Essays from Thirty Years. *New Edn. Routledge* (December 20, 1995) (London).

Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Front. Comput. Neurosci.* 6:33. doi: 10.3389/fncom.2012.00033

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

Priem, J., and Hemminger, B. M. (2012). Decoupling the scholarly journal. *Front. Comput. Neurosci.* 6:19. doi: 10.3389/fncom.2012.00019

Priem, J., Piwowar, H. A., and Hemminger, B. M. (2012). Altmetrics in the wild: using social media to *explore scholarly impact.* ArXiv 1203.4745

Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Disc.* 10, 712. doi: 10.1038/nrd3439-c1

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., and Stirling, A. (2012). How journal rankings

can suppress interdisciplinary research: a comparison between Innovation Studies and Business and Management. *Res. Policy* 41, 1262–1282. doi: 10.1016/j.respol.2012.03.015

Reedijk, J. (1998). Sense and nonsense of science citation analyses: comments on the monopoly position of ISI and citation inaccuracies. *Risks of possible misuse and biased citation and impact data. New J. Chem.* 22, 767–770. doi: 10.1039/a802808g

Research Information Network. (2008). Activities, Costs and Funding Flows in the Scholarly Communications System *Research Information Network*. Report commissioned by the Research Information Network (RIN). Available online at: http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu

Roemer, R. C., and Borchardt, R. (2012). From bibliometrics to alt-metrics: a changing scholarly landscape. *College Res. Libr. News* 73, 596–600.

Rossner, M., Van Epps, H., and Hill, E. (2007). Show me the data. *J. Cell Biol.* 179, 1091–1092. doi: 10.1083/jcb.200711140

Saha, S., Saint, S., and Christakis, D. A. (2003). Impact factor: a valid measure of journal quality? *JMLA* 91, 42–46.

Sandewall, E. (2012). Maintaining live discussion in two-stage open peer review. *Front. Comput. Neurosci.* 6:9. doi: 10.3389/fncom.2012.00009

Sarewitz, D. (2012). Beware the creeping cracks of bias. *Nature* 485, 149–149. Available online at: http://www.nature.com/doifinder/10.1038/485149a [Accessed May 10, 2012]. doi: 10.1038/485149a

Sarewitz, D. (2013). Science must be seen to bridge the political divide. *Nature* 493:7. doi: 10.1038/493007a

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature* 470, 437. doi: 10.1038/470437a

Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., et al. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph. Lateral Scler.* 9, 4–15. doi: 10.1080/17482960701856300

Seglen, P. O. (1992). The skewness of science. *J. Am. Soc. Inf. Sci.* 43, 628–638. doi: 10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0

Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. *J. Am. Soc. Inf. Sci.* 45, 1–11. doi: 10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ* 314, 498–502. doi: 10.1136/bmj.314.7079.497

Shapin, S., (2008). *The Scientific Life?: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press. doi: 10.7208/chicago/9780226750170.001.0001

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). Falsepositive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Simmons, L. W., Tomkins, J. L., Kotiaho, J. S., and Hunt, J. (1999). Fluctuating paradigm. *Proc. R. Soc. B Biol. Sci.* 266, 593–595. doi: 10.1098/rspb.1999.0677

Singh, G., Haddad, K. M., and Chow, C. W. (2007). Are articles in "Top" management journals necessarily of higher quality? *J. Manag. Inquiry* 16, 319–331. doi: 10.1177/1056492607305894

Siontis, K. C. M., Evangelou, E., and Ioannidis, J. P. A. (2011). Magnitude of effects in clinical trials published in high-impact general medical journals. *Int. J. Epidemiol.* 40, 1280–1291. doi: 10.1093/ije/dyr095

Sønderstrup-Andersen, E. M., and Sønderstrup-Andersen, H. H. K. (2008). An investigation into diabetes researcher's perceptions of the Journal Impact Factor—reconsidering evaluating research. *Scientometrics* 76, 391–406. doi: 10.1007/s11192-007-1924-4

Song, F., Eastwood, A., Gilbody, S., and Duley, L. (1999). The role of electronic journals in reducing publication bias. *Inf. Health Soc. Care* 24, 223–229. doi: 10.1080/146392399298429

Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Org. Sci.* 16, 180–200. doi: 10.1287/orsc.1040.0107

Statzner, B., and Resh, V. H. (2010). Negative changes in the scientific publication process in ecology: potential causes and consequences. *Freshw. Biol.* 55, 2639–2653. doi: 10.1111/j.1365-2427.2010.02484.x

Steen, R. G. (2011a). Retractions in the scientific literature: do authors deliberately commit research fraud? *J. Med. Ethics* 37, 113–117. doi: 10.1136/jme.2010.038125

Steen, R. G. (2011b). Retractions in the scientific literature: is the incidence of research fraud increasing? *J. Med. Ethics* 37, 249–253. doi: 10.1136/jme.2010.040923

Sutton, J. (2011). psi study highlights replication problems. *The Psychologist News*. Available online at: http://www.thepsychologist.org.uk/blog/blogpost.cfm?threadid=1984andcatid=48

Taylor, M., Perakakis, P., and Trachana, V. (2008). The siege of science. *Ethics Sci. Environ. Polit.* 8, 17–40. doi: 10.3354/esep00086

Todd, P., and Ladle, R. (2008). Hidden dangers of a 'citation culture'. *Ethics Sci. Environ. Polit.* 8, 13–16. doi: 10.3354/esep00091

Tressoldi, P. E., Giofré, D., Sella, F., and Cumming, G. (2013). High impact = high statistical standards? Not necessarily so. *PLoS ONE* 8:e56180. doi: 10.1371/journal.pone.0056180

Tsikliras, A. (2008). Chasing after the high impact. *Ethics Sci. Environ. Polit.* 8, 45–47. doi: 10.3354/esep00087

Upadhyay, J., Baker, S. J., Chandran, P., Miller, L., Lee, Y., Marek, G. J., et al. (2011). Default-mode-like network activation in awake rodents. *PLoS ONE* 6:e27839. doi: 10.1371/journal.pone.0027839

Van Dongen, S. (2011). Associations between asymmetry and human attractiveness: possible direct effects of asymmetry and signatures of publication bias. *Ann. Hum. Biol.* 38, 317–323. doi: 10.3109/03014460.2010.544676

Van Noorden, R. (2011). Science publishing: the trouble with retractions. *Nature* 478, 26–28. doi: 10.1038/478026a

Van Noorden, R. (2013). Mathematicians aim to take publishers out of publishing. *Nat. News*. Available online at: http://www.nature.com/news/mathematicians-aim-to-take-publishers-out-of-publishing-1.12243

Vanclay, J. K. (2011). Impact factor: outdated artefact or stepping-stone to journal certification? *Scientometrics* 92, 211–238. doi: 10.1007/s11192-011-0561-0

Wager, E., and Williams, P. (2011). Why and how do journals retract articles? An analysis of Medline retractions 1988–2008. *J. Med. Ethics* 37, 567–570. doi: 10.1136/jme.2010.040964

Wagner, P. D. (2011). What's in a number? *J. Appl. Physiol.* (Bethesda, MD.: 1985) 111, 951–953. doi: 10.1152/japplphysiol.00935.2011

Walther, A., and Van den Bosch, J. J. F. (2012). FOSE: a framework for open science evaluation. *Front. Comput. Neurosci.* 6:32. doi: 10.3389/fncom.2012.00032

Weale, A. R., Bailey, M., and Lear, P. A. (2004). The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. *BMC Med. Res. Methodol.* 4:14. doi: 10.1186/1471-2288-4-14

Welberg, L. (2012). Neuroimaging: rats join the "default mode" club. *Nat. Rev. Neurosci.* 11:223. doi: 10.1038/nrn3224

Wicherts, J. M., Kievit, R. A., Bakker, M., and Borsboom, D. (2012). Letting the daylight in: reviewing the reviewers and other ways to maximize transparency in science. *Front. Comput. Neurosci.* 6:20. doi: 10.3389/fncom.2012.00020

Yarkoni, T. (2012). Designing next-generation platforms for evaluating scientific output: what scientists can learn from the social web. *Front. Comput. Neurosci.* 6:72. doi: 10.3389/fncom.2012.00072

Yong, E. (2012). Replication studies: bad copy. *Nature* 485, 298–300. doi: 10.1038/485298a

Young, N. S., Ioannidis, J. P. A., and Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Med.* 5:e201. doi: 10.1371/journal.pmed.0050201

Yue, W., Wilson, C. S., and Boller, F. (2007). Peer assessment of journal quality in clinical neurology. *JMLA* 95, 70–76.

Zimmermann, J., Roebroeck, A., Uludag, K., Sack, A., Formisano, E., Jansma, B., et al. (2011). Network-based statistics for a community driven transparent publication process. *Front. Comput. Neurosci.* 6:11. doi: 10.3389/fncom.2012.00011

APPENDIX

FIGURE A1 | Impact Factor of the journal "Current Biology" in the years 2002 (above) and 2003 (below) showing a 40% increase in impact. The increase in the IF of the journal "Current Biology" from approx. 7 to almost 12 from one edition of Thomson Reuters' "Journal Citation Reports" to the next is due to a retrospective adjustment of the number of items published (marked), while the actual citations remained relatively constant.