
Link to official URL (if available):

Opus: University of Bath Online Publication Store

http://opus.bath.ac.uk/

This version is made available in accordance with publisher policies. Please cite only the published version using the reference above.

See http://opus.bath.ac.uk/ for usage policies.

Please scroll down to view the document.
Deep impact: unintended consequences of journal rank

Björn Brembs1*, Katherine Button2 and Marcus Munafò3

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

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.

RETRACEMENTS 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
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 Munafò, 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).
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 signals which either failed to replicate the original findings, or journals that 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 the fact that journal rank is a poor predictor of statistical soundness is the fact 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 = −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...
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
there is quite extensive evidence arguing for the strong correlation due to the influence of the IF on reading habits (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) 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 (Upadhayay 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 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

---

**Table 1 | Thomson Reuters’ IF calculations for the journal “Current Biology” in the years 2002/2003.**

<table>
<thead>
<tr>
<th>Year</th>
<th>Published Items 2000</th>
<th>Published Items 2001</th>
<th>Published Items 2002</th>
<th>Published Items 2003</th>
<th>Citations in preceding 2 years</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCR science edition 2002</td>
<td>504</td>
<td>528</td>
<td>n.c.</td>
<td>1032</td>
<td>7231</td>
<td>7.007</td>
</tr>
<tr>
<td>JCR science edition 2003</td>
<td>n.c.</td>
<td>300</td>
<td>334</td>
<td>634</td>
<td>7551</td>
<td>11.910</td>
</tr>
</tbody>
</table>

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.
conclusions, but do emphasize the need for future research. 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; Sonderstrup-Andersen and Sonderstrup-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 inverting 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 noth-
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 (Hönkopp 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; Jutro-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.

**ACKNOWLEDGMENTS**

Neil Saunders was of tremendous value in helping us obtain and understand the PubMed retraction data for Figure 1A. Ferric Fang and Arturo Casadeval were so kind as to let us use their retraction data to re-plot their figure on a logarithmic scale (Figure 1D). We are grateful to George A. Lozano, Vincent Larivière, and Yves Gingras for sharing their citation data with us (Figure 3). We are indebted to John Ioannidis, Daniele Fanelli, Christopher Baker, Dwight Kravitz, Tom Hartley, Jason Priem, Stephen Curry and three anonymous reviewers for their comments on an earlier version of this manuscript, as well as to Nikolaus Kriegeskorte and Erin Warnsley for their helpful comments and suggestions during the review process here. Marcus R. Munafò is a member of the UK Centre for Tobacco Control Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. Björn Brembs was a Heisenberg-Fellow of the German Research Council during the time most of this manuscript was written and their support is gratefully acknowledged as well.


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 January 2013; accepted: 03 June 2013; published online: 24 June 2013.


Copyright © 2013 Brembs, Button and Munafò. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided that the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.
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.