The Impact Factor Fallacy

Frieder M. Paulus[1]*, Nicole Cruz[2,3]* and Sören Krach[1]*

[1] *Department of Psychiatry and Psychotherapy, Social Neuroscience Lab, University of Lübeck, Lübeck, Germany,*
[2] *Department of Psychological Sciences, Birkbeck, University of London, London, United Kingdom,* [3] *Laboratoire CHArt, École Pratique des Hautes Études, Paris, France*

The use of the journal impact factor (JIF) as a measure for the quality of individual manuscripts and the merits of scientists has faced significant criticism in recent years. We add to the current criticism in arguing that such an application of the JIF in policy and decision making in academia is based on false beliefs and unwarranted inferences. To approach the problem, we use principles of deductive and inductive reasoning to illustrate the fallacies that are inherent to using journal-based metrics for evaluating the work of scientists. In doing so, we elaborate that if we judge scientific quality based on the JIF or other journal-based metrics we are either guided by invalid or weak arguments or in fact consider our uncertainty about the quality of the work and not the quality itself.

Keywords: journal impact factor, journal ranking, scientific excellence, open access, publishing, scholarly communication, reasoning

INTRODUCTION

The journal impact factor (JIF) was initially used to help librarians make decisions about journals (Garfield, 2006). However, during the last decades the usage of the JIF has significantly changed. In deviating from its original purpose it is now widely used to evaluate the quality of individual publications and the work of scientists (Amin and Mabe, 2003; Arnold and Fowler, 2010). Since then, the measure itself has been extensively criticized for various reasons. For example, it is well known that the JIF is an inaccurate estimate for the expected number of citations of an article within a specific journal (Callaway, 2016; Larivière et al., 2016), that it is relatively easy to manipulate (McVeigh and Mann, 2009; Tort et al., 2012), it does not show noticeable positive associations with objective measures of the quality of the science (Yeung, 2017; Brembs, 2018), and might be affected by the increasing usage of preprint servers[1]. While some bibliometricians aim to establish distinctive and perhaps more justified use cases of the JIF in approximating the visibility of publications instead of using it as a measure of quality (Gumpenberger et al., 2012; Gorraiz et al., 2016, 2017), the JIF nonetheless has deeply affected the work of scientists and decision making in academia as a measure for scientific excellence. Scientists get jobs, tenure, grants, and bonuses based on the impact of the journals they are publishing their manuscripts in, outgrowths whose consequences were critically discussed in many previous reviews, comments and editorials (Seglen, 1997; Lehmann et al., 2006; Della Sala and Crawford, 2007; Simons, 2008; Reich, 2013; Casadevall and Fang, 2014; DePellegrin and Johnston, 2015; Werner, 2015; Stephan et al., 2017 and please see Brembs et al., 2013 for very thorough analyses of the detrimental effects of measures

---

[1]A further, and less frequently discussed bias might be introduced through the increasing usage of preprint servers. Here, articles which have been published as preprints continue to gather citations even if the article, or a revised version of the articles has been published in a journal. The citations of the preprints are yet not added to the citation count of the subsequently published version of the article (Davis, 2018). This opens the door for a variety of novel biases for the JIF e.g., if authors are more likely to update their citation library in case of preprints being published in journals with higher JIF compared to preprints published in journals with lower JIF.

of journal rank). Notably, the JIF has also been explicitly referred to as a tool to decide how to distribute funds across institutions, for example in Germany (Deutsche Forschungsgemeinschaft (DFG) [German Science Foundation], 2004), and thereby affects policy making on a much larger scale.

"For the calculation of the performance-based bonus of the unit providing the service (department or clinic) the original publications may be used with the unweighted impact factor of the publication organ, in the sense of a step-wise introduction of quality criteria. Thereby, a first- and last authorship may be considered with one third each and the remaining third can be distributed across all remaining authors [...]."[2]

Besides such explicit usage of the JIF for evaluating scientific excellence, the JIF also implicitly affects other measures which have been suggested to better approximate the quality of a scientist's work or of a specific study [e.g., the h-index, Hirsch, 2005 and the Relative Citation Ratio (RCR), Hutchins et al., 2015]. For example, there is some evidence that the number of citations of an article is influenced by the JIF of the journal where the article was published, regardless of the quality of the article itself (Callaham et al., 2002; Lozano et al., 2012; Brembs et al., 2013; Cantrill, 2016). This implies that measures that are based on the citations of the individual articles are still influenced by the JIF of the publication organ. With the many different ways of how the JIF can influence decision making in academia and is used as an explicit tool for policy making and staffing[3], it is not surprising that empirical data now demonstrate the JIF to be one of the most powerful predictors for academic success (Van Dijk et al., 2014). Accordingly, we could recently show that some scientists may have adapted to these reward principles in their environment by showing a greater reward signal in the brain's reward structures in the prospect of an own high impact publication (Paulus et al., 2015).

In line with the rising initiatives to prevent the use of the JIF for evaluating the quality of science (see e.g., the DORA initiative, Alberts, 2013; Cagan, 2013 or the report of the Wissenschaftsrat [German Council of Science and Humanities], 2015), we have considerable doubts that the arguments in

support of using the JIF for measuring scientific excellence are justified. In this comment we want to look at the problem of using the JIF from a different perspective and carefully (re)evaluate the arguments for its use as an estimate of scientific quality based on principles of deductive and inductive reasoning. Thereby, we hope to better understand the beliefs about the JIF that influence decisions in academia and the implications of policies that use the JIF to assess and remunerate scientific quality. Beyond the specific case of the JIF, this exercise might also help to specify more general misconceptions when using measures for journal rank to evaluate science, in order to overcome incentive structures based on journal-based metrics altogether.

DEDUCTIVE FALLACY WHEN USING THE JIF

A basic belief when using the JIF for evaluating the quality of a specific manuscript seems to be that (1) if a paper is published in a high impact factor journal (p) then the paper is of high quality (q)[4]. Why would scientists believe this? A straightforward reason is the idea that it is more difficult to publish in a high impact factor journal because higher standards of research quality and novelty have to be passed in order to be accepted. The average number of citations of a journal's articles within in a specific time period signals the average breadth of interest in these articles during that time period, which can of course be affected by many factors other than research quality. But as a first approximation, let us suppose that belief (1) is the case. What can we conclude from it? If we see a paper published in a high impact factor journal, we could then draw the deductively valid inference of modus ponens (MP: *if p then q, p, therefore q*)[5] and conclude that the paper is of high quality. But what if we see a paper published in a low impact factor journal? Can we draw any conclusions in this case?

One aspect of the impact factor fallacy could be operationalized as the tendency to draw the deductively invalid inference of *denial of the antecedent* (DA: *if p then q, not-p, therefore not-q*). This inference is deductively invalid because it is logically consistent for the premises *if p then q* and *not-p* to be true and yet the conclusion *not-q* to be false. When the premises of an inference can be true and at the same time the conclusion false, the inference does not preserve truth when going from premises to conclusion. In order to argue that the conclusion is not false in a particular case, we would therefore have to go beyond this argument and provide further information that might increase support for the conclusion.

For the more realistic case that the premises and conclusion are uncertain, such that they can not only be either true

---

[3]During the last 4 years, two of the authors (FMP, SK) have noted several calls for senior positions in medical departments where applicants were formally required to provide JIF based information such as the accumulated JIF or average JIF. While we do not have information about the generalizability of these practices we think it is not surprising that institutions who distribute funds according to JIF also select people based on this measure following the idea that "the best predictor for future behavior is past behavior." Notably, the circularity – in the sense that people are selected for senior positions based on the JIF and the JIF is then used as a predictor for career success – might very well contribute to the association between the average JIF of a scientists' publications and their success on the academic job market (Van Dijk et al., 2014). This association might thus rather reflect staffing policies instead of providing evidence for the usefulness of the JIF as a predictor of academic achievements.

[4]When we speak of "high" and "low" impact in this manuscript, the arguments we make are independent of whether "high" and "low" refer to the absolute JIF of a journal, or to the JIF relative to a specific research domain.

[5]Here *p* and *q* stand for arbitrary propositions. For example, *p* might stand for "This paper is published in a high impact factor journal" and *q* for "This paper is of high quality."

---

or false, but can be held with varying degrees of belief, the inference of DA is probabilistically invalid (*p*-invalid) because there are coherent[6] probability assignments to premises and conclusion for which the probability of the conclusion is lower than the sum of the probabilities of the premises (Adams, 1998; Over, 2016). The probabilistic case is therefore analogous to the binary one. In the binary case, DA does not preserve truth from premises to conclusion; and in the probabilistic case, DA does not preserve probability from premises to conclusion. The latter means that it can be warranted to have a high degree of belief in the premises and yet a very low degree of belief in the conclusion. In order to justify the conclusion in a particular instantiation of the argument, we would have to bring further information into the discussion beyond that contained in the premises. Applied to the JIF example, suppose we assume that if a paper is published in a high impact factor journal, it is of high quality, and then encounter a paper that is published in a low impact factor journal. From this alone it is not justified to conclude that the paper we encountered is not of high quality[7]. In order to draw such a conclusion, we would require more information.

Denial of the antecedent (DA) is of course not the only inference one can draw based on the conditional belief that if a paper is published in a high impact factor journal, then it is of high quality. A similar, deductively valid inference results if we add a further premise to DA: "If a paper is not published in a high impact factor journal, then it is not of high quality." One can combine this new conditional premise with the conditional premise that we already had: "If a paper is published in a high impact factor journal, then it is of high quality," to obtain the following biconditional premise: "A paper is published in a high impact factor journal if and only if it is of high quality." From this biconditional premise (or equivalently from the two conditional premises) together with the premise that a specific paper was not published in a high impact factor journal, one can indeed validly conclude that the paper is not of high quality. However, this inference will only be useful if one believes the biconditional premise to a non-negligible degree in the first place. If the biconditional premise is implausible, then any deductively valid conclusion based on it will also tend to be implausible, precisely because it follows logically from an implausible starting assumption. Considering that most scientists are likely to agree that it is not only implausible but false that a paper is of high quality if and only if it is

published in a high impact factor journal, the fact that the inference from this biconditional is valid has no use for practical purposes.

INDUCTIVE FALLACIES WHEN USING THE JIF

One could argue that deduction, and with it logical validity, has little impact on actual reasoning and decision making outside of the mathematics classroom, and that therefore the inferences we should be looking at when analyzing the use of the JIF in the practice of science should rather be inductive (Evans, 2002; Oaksford and Hahn, 2007; Chater et al., 2011; Baratgin and Politzer, 2016).

An inductive inference that might describe well the use of the impact factor is the informal fallacy of the *argument from ignorance* (or its Latin equivalent "ad ignorantiam"). This argument tries to justify a conclusion by pointing out that there is no evidence against it. Typical examples could be "No side effects were found for this treatment in clinical trials. Therefore this treatment is safe" or "No one has proven that ghosts do not exist. Therefore ghosts exist" (Oaksford and Hahn, 2004, 2007; Hahn and Oaksford, 2007). In the case of the JIF, if a paper comes from a high impact journal this can be seen as a sign suggesting it is an excellent piece of work. But as we saw above in the discussion of DA, this does not imply that if the paper was published in a low impact factor journal this is a sign suggesting that the quality of the paper is low. A more precise description of the situation would be that a low impact factor journal lacks the sign of high quality that a high JIF provides. If a paper is published in a low impact journal then we have less information about its quality, rather than having information suggesting that its quality is low. It is an argument from ignorance to use the absence of impact factor based evidence for high quality to conclude that a paper is of low quality.

However, the argument from ignorance is not always a bad argument (Hahn and Oaksford, 2007, 2012). Its strength depends on how informative the lack of information about something being the case is in the situation at hand. Suppose we search a book in a library catalog and do not find it. In this case it is reasonable to use the lack of information about the book in the catalog to conclude that the book is not in the library. Similarly, if we look at a train timetable and do not see a particular town listed, it is reasonable to conclude that the train does not stop in that town. However, suppose we are planning a party and have invited the whole department, in the hope that a particular person we are attracted to will attend. In this case a lack of information indicating that the person will come does not warrant the conclusion that the person will not come. Catalogs and timetables are fairly closed environments in which we can expect all relevant information to be stated explicitly. But environments like those of social interactions or research endeavors with myriads of available scholarly communication channels are more open, so that the absence of information about something being the case simply does not warrant us to conclude that it is not the case. A consequence for the JIF would be that

---

[6]Two statements are *coherent* if and only if they respect the axioms of probability theory. For example, these axioms state that if we believe it is 80% likely to rain, then in order for our beliefs to be coherent we should also be willing to believe that it is 20% likely not to rain, otherwise the probabilities involved would not sum up to 1.

[7]Notably, this argument also applies to manuscripts uploaded on preprint servers which have not been peer-reviewed. In line of this argument, here the lack of peer-review does not necessarily imply inferior quality. We think that preprint repositories are very interesting testbeds, since the fundamental lack of information on the quality of articles through journal-based metrics (see the section on inductive inference below) might motivate scientists and institutions to evaluate the quality and importance of the work more by reading the articles instead of relying on numbers or reputation of publication organs.

low impact publications do not signal low research quality, but rather uncertainty about the quality and the need to gather more information in order to be able to determine research quality.

Two further inductive inferences that might be relevant in accounting for the use of the JIF are the informal fallacies of the *argument from authority* (also called by the Latin name "ad verecundiam"), and of the *ad hominem argument* (Bhatia and Oaksford, 2015; Hahn and Hornikx, 2016). The argument from authority tries to justify a conclusion by pointing out that some expert or authority endorses the conclusion. Typical examples could be "Scientist x says that the treatment is safe. Therefore the treatment is safe," "My parents say that Santa Claus exists. Therefore Santa Claus exists" or "My peers say that clothing item x is great. Therefore clothing item x is great." In the case of the JIF, a high impact factor of a journal would play the role of an authority for the quality of the papers within it.

In contrast, the ad hominem argument tries to justify the rejection of a conclusion by pointing to personal attributes of a person that endorses it. Typical examples could be "The new treatment was developed by a person with no formal degree in the subject. Therefore the treatment is not safe," or "A person without a driver's license says "don't drink alcohol while driving." Therefore, it is false that you should not drink alcohol while driving." In the case of the JIF, a low impact factor would be used to give a journal a reputation of low quality, and this low quality reputation would then be transferred to the papers within it.

Like the argument from ignorance, the argument from expert opinion and the ad hominem argument are not always bad arguments. Their quality varies as a function of how informative the authority status, or the personal attributes of the instance endorsing them, is for the problem at hand. Policy decisions are routinely based on the advice of experts, and there seems to be agreement that this is a good thing to do, as long as the experts are really considered experts in their field and their advice is not biased (Harris et al., 2016; c.f. Sloman and Fernbach, 2017). Dismissing an argument because of personal attributes of a person endorsing it is often more difficult, because it has to be made plausible that those attributes are relevant to the quality of the argument. For example, that one does not need to be a mother to be qualified for being prime minister seems obvious, whereas a case of a person applying to a position against gender discrimination, who in his private life beats his wife, is likely to be more controversial. In the case of the JIF, we would have to justify why we think that a low impact factor indicates that a particular journal is of low quality, and why this low quality can be transferred to a particular paper within it. Such a judgment requires further information about the journal and about the paper at hand to be justified, which is usually not provided and difficult to obtain since it might e.g., not be clear if review processes in lower JIF journals are less able to detect errors (Brembs, 2018). Thus, whereas a high impact factor may add to the reputation of a journal, a low impact factor does not warrant a bad reputation, but rather provides insufficient information about reputation (see **Table 1** for examples of the inductive and deductive fallacies as discussed here).

Until now we have discussed inferences on the basis of the belief that if a paper is published in a high impact factor journal, then it is of high quality. But although this belief can sometimes be useful as a quick approximation or rule of thumb, it is often itself not warranted.

First, the mean number of citations of the papers in a journal reflect of the average breadth of interest in these papers during the first years after publication, which is not the same as research quality. For example, a high quality paper may have a low citation rate because research in specific fields (e.g., the humanities) is less dependent on the cumulative knowledge of what goes before (Van Noorden, 2017), or because it is going unnoticed for a long time (a so-called "sleeping beauty") and only "awakened by a prince" years after publication (c.f. van Raan, 2004). Conversely, a paper may also have high citation rates because of highly consequential flaws within it, which attracts attention and stimulates the scientific discourse.

Secondly, it is often not warranted because the inference from a metric defined at the journal level to the features of an individual paper within that journal might involve an ecological fallacy.

ECOLOGICAL FALLACY WHEN USING THE JIF

Finally, the evaluation of manuscripts based on the JIF bears an ecological fallacy. When comparing group level data, such as the average citations of journals, it is difficult up to impossible to infer the likelihood of the outcome for comparisons on the individual level, such as citations of manuscripts. In fact, it is relatively easy to think of examples where the likelihood to find a manuscript with more than 12 citations per year in a lower impact journal exceeds the likelihood of finding such manuscript in a higher impact journal. This type of ecological fallacy occurs when the distribution of citations is heavily and differentially skewed within each higher-level unit, i.e., the journals. This is typically the case when it comes to citation rates of journals (see e.g., Weale et al., 2004; Larivière et al., 2016).

Accordingly, a journal with a JIF of 12 might contain few manuscripts that were cited several hundred times in the previous 2 years, but many others that were rarely cited during the same period. Such a citation pattern would result in a heavily skewed distribution of citations per article, while another journal with a JIF of 10 might have a normally distributed citation rate of articles for the same time period. Without further knowledge of the distribution of citations within the journals in a given year (i.e., information at the individual level) concluding that a manuscript in the journal with a higher JIF is of better quality (or of broader interest) might bear an ecological fallacy, because it is possible that the likelihood of finding a manuscript with more citations in the lower impact journal is not reflected by the JIF.

CONCLUDING REMARKS

With this comment, we hope to have highlighted some misconceptions in the beliefs and arguments involved in using

TABLE 1 | The deductive and inductive fallacies discussed in this paper.

Name	Form	Plausible example	Implausible example	Journal impact factor example
Deductive fallacy				
Denial of the antecedent	*If p then q. Not-p.Therefore not-q.*	If the glass falls down then it breaks. The glass does not fall down. Therefore, the glass does not break.	If you carry an umbrella then you stay dry. You do not carry an umbrella. Therefore, you do not stay dry.	If a paper is published in a high impact factor journal, then it is of high quality. This paper is not published in a high impact factor journal. Therefore, this paper is not of high quality.
Inductive fallacies				
Argument from ignorance	*It is not known that p is true (false). Therefore p is false (true).*	The book is not listed in the library catalog. Therefore, the book is not in the library.	No one has proven that ghosts do not exist. Therefore, ghosts exist.	This paper does not have the quality sign of having been published in a high impact factor journal. Therefore, this paper is not of high quality.
Argument From authority	*This expert says that p is true. Therefore p is true.*	Medical experts say that this treatment is safe. Therefore, this treatment is safe.	My parents say that Santa Claus exists. Therefore, Santa Claus exists.	This paper does not have the authority backing of having been published in a high impact factor journal. Therefore, this paper is not of high quality.
Ad hominem argument	*This untrustworthy person says that p is true. Therefore p is false.*	A person without training says that this treatment is safe. Therefore, this treatment is not safe.	A person without a driver's license says "don't drink alcohol while driving." Therefore, it is false that you should not drink alcohol while driving.	This paper was published in a journal with low quality reputation due to a low impact factor. Therefore, this paper is not of high quality.

journal-based metrics, and specifically the JIF, for evaluating the work of scientists. While some of the thoughts described here are introduced to illustrate the most controversial arguments, others better approximate the reality of decision making in academia. In this exercise, we think it became clear that without the specification of further premises besides the simple idea that publishing in high-impact journals is an indicator for high scientific quality the validity of the conclusions is not warranted. These additional premises need to be explicitly expressed in the argument, otherwise the plausibility of the premises and accordingly the conclusion cannot be verified. Only if one believes in the premises to a non-negligible degree in first place, for example that manuscripts published in journals with a low JIF are of low quality, this could end up in a plausible conclusion in case of a valid deduction. Similarly, also for inductive inferences additional explicit assumptions are needed to arrive at plausible conclusions. The simple lack of a journal-based quality sign is not sufficient to conclude that a manuscript is of low quality. One could consider scientific publishing a closed environment where people follow similar rules and have no degrees of freedom regarding where and why to submit their work to provide a strong argument for this case. Given the complexity of the matter, it is surprising to see many political and academic institutions as well as scientists argue that they are evaluating the "quality of science" while providing weak arguments, drawing invalid conclusions, or weighing their lack of information and uncertainty about the subject when using the JIF.

From an economic perspective, however, it might in fact be a successful strategy to minimize the uncertainty about the quality of the evaluated work, person, or institution by relying on the JIF or other journal based metrics (Nature Editorial, 2018), and it might also be better to have a weak argument than to have no argument. Evaluating the quality of a scientist's work surely is a time-consuming process and it takes much more effort than simply comparing impact factors. Accordingly, deans, commissions, or institutions which might not have the resources for an actual assessment of "scientific excellence" have reasons to rely on the JIF. However, it should be clear that those decisions are not based on the *quality* of the scientific contribution *per se* but, optimistically, somehow integrate the *availability of information* about the quality. This distinction makes an important difference for communicating and justifying decisions in academia. As an illustrative example, one can compare the situation of deciding that a candidate does not deserve tenure because one thinks that the quality of the work was not good enough, to deciding that a candidate does not deserve tenure because one lacks information and is uncertain whether the quality of the work was good enough. While persons and institutions usually *communicate* as if they were following the first argument, their *justification* most often implies the latter if they base their decisions on journal-based metrics.

The JIF is arguably the most popular journal-based metric of our times, but it has already been subject to severe criticism in the past (Seglen, 1997; Lehmann et al., 2006; Della Sala and Crawford, 2007; Simons, 2008; Brembs et al., 2013; Reich, 2013; DePellegrin and Johnston, 2015; Werner, 2015). As a result, it seems that (some) individuals and institutions within the scientific community are ready to shake off the JIF at some point in the nearer future (Alberts, 2013; Cagan, 2013;

Callaway, 2016). We want to point out that the problems described here apply in one way or another to any journal-based assessment. If journals would drop out of the "impact factor game" (PLoS Medicine Editorial, 2006) publications in some journals might still be regarded as more valuable than in others. It is difficult to quantify those influences, but having a publication in one of the "golden club" journals (Reich, 2013) could simply replace the metric of the JIF with another, more implicit qualitative measure for distinguishing prestigious from less prestigious journals. Thereby, the fallacies and problems described above would continue to govern decision making in academia as long as we base them on any kind of journal-based assessment and the rank of publication organs.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Adams, E. W. (1998). *A Primer of Probability Logic*. Stanford, CA: CSLI.

Alberts, B. (2013). Impact factor distortions. *Science* 340:787. doi: 10.1126/science.1240319

Amin, M., and Mabe, M. A. (2003). Impact factors: use and abuse. *Medicina* 63, 347–354.

Arnold, D. N., and Fowler, K. K. (2010). *Nefarious Numbers*. Availabe at: http://arxiv.org/abs/1010.0278

Baratgin, J., and Politzer, G. (2016). "Logic, probability and inference: a methodology for a new paradigm," in *Cognitive Unconscious and Human Rationality*, eds L. Macchi, M. Bagassi, and R. Viale (Cambridge, MA: MIT Press).

Bhatia, J.-S., and Oaksford, M. (2015). Discounting testimony with the argument ad hominem and a Bayesian congruent prior model. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1548–1559. doi: 10.1037/xlm0000151

Brembs, B. (2018). Prestigious science journals struggle to reach even average reliability. *Front. Hum. Neurosci.* 12:37. doi: 10.3389/fnhum.2018.00037

Brembs, B., Button, K., and Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Front. Hum. Neurosci.* 7:291. doi: 10.3389/fnhum.2013.00291

Cagan, R. (2013). The San Francisco declaration on research assessment. *Dis. Models Mech.* 6, 869–870. doi: 10.1242/dmm.012955

Callaham, M., Wears, R. L., and Weber, E. (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

Callaway, E. (2016). Beat it, impact factor! Publishing elite turns against controversial metric. *Nature* 535, 210–211. doi: 10.1038/nature.2016.20224

Cantrill, S. (2016). *Imperfect Impact*. Availabe at: https://stuartcantrill.com/2016/01/23/imperfect-impact/

Casadevall, A., and Fang, F. C. (2014). Causes for the persistence of impact factor mania. *mBio* 5, e01342-14. doi: 10.1128/mBio.00064-14

Chater, N., Oaksford, M., Hahn, U., and Heit, E. (2011). "Inductive logic and empirical psychology," in *Handbook of the History of Logic*, Vol. 10, eds D. M. Gabbay and J. Woods (Amsterdam: Inductive Logic), 553–624.

Davis, P. (2018). *Journals Lose Citations to Preprint Servers*. Available at: https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories/

Della Sala, S., and Crawford, J. R. (2007). A double dissociation between impact factor and cited half life. *Cortex* 43, 174–175. doi: 10.1016/S0010-9452(08)70473-8

DePellegrin, T. A., and Johnston, M. (2015). An arbitrary line in the sand: rising scientists confront the impact factor. *Genetics* 201, 811–813. doi: 10.1534/genetics.115.182261

Deutsche Forschungsgemeinschaft (DFG) [German Science Foundation] (2004). *Empfehlungen zu einer »Leistungsorientierten Mittelvergabe« (LOM) an den Medizinischen Fakultäten: Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft*. Bonn: Deutsche Forschungsgemeinschaft.

Evans, J. S. (2002). Logic and human reasoning: an assessment of the deduction paradigm. *Psychol. Bull.* 128, 978–996. doi: 10.1037/0033-2909.128.6.978

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA* 295, 90–93. doi: 10.1001/jama.295.1.90

Gorraiz, J., Wieland, M., and Gumpenberger, C. (2016). Individual bibliometric assessment at University of Vienna: from numbers to multidimensional profiles. *Prof. Inf.* 25:901. doi: 10.3145/epi.2016.nov.07

Gorraiz, J., Wieland, M., and Gumpenberger, C. (2017). To be visible, or not to be, that is the question. *Int. J. Soc. Sci. Hum.* 7, 467–471.

Gumpenberger, C., Wieland, M., and Gorraiz, J. (2012). Bibliometric practices and activities at the University of Vienna. *Libr. Manage.* 33, 174–183. doi: 10.1108/01435121211217199

Hahn, U., and Hornikx, J. (2016). A normative framework for argument quality: argumentation schemes with a Bayesian foundation. *Synthese* 193, 1833–1873. doi: 10.1007/s11229-015-0815-0

Hahn, U., and Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychol. Rev.* 114, 704–732. doi: 10.1037/0033-295X.114.3.704

Hahn, U., and Oaksford, M. (2012). "Rational argument," *The Oxford Handbook of Thinking and Reasoning*, eds K. J. Holyoak and R. G. Morrison (New York, NY: Oxford University Press). doi: 10.1093/oxfordhb/9780199734689.013.0015

Harris, A. J. L., Hahn, U., Madsen, J. K., and Hsu, A. S. (2016). The appeal to expert opinion: quantitative support for a Bayesian network approach. *Cogn. Sci.* 40, 1496–1533. doi: 10.1111/cogs.12276

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16569–16572. doi: 10.1073/pnas.0507655102

Hutchins, B. I., Yuan, X., Anderson, J. M., and Santangelo, G. M. (2015). Relative Citation Ratio (RCR): a new metric that uses citation rates to measure influence at the article level. *PLoS Biol.* 14:e1002541. doi: 10.1371/journal.pbio.1002541

Larivière, V., Kiermer, V., Maccallum, C. J., Mcnutt, M., Patterson, M., Pulverer, B., et al. (2016). A simple proposal for the publication of journal citation distributions. *BioRxiv* [Preprint]. doi: 10.1101/062109

Lehmann, S., Jackson, A. D., and Lautrup, B. E. (2006). Measures for measures. *Nature* 444, 1003–1004. doi: 10.1038/4441003a

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

McVeigh, M. E., and Mann, S. J. (2009). The journal impact factor denominator. *JAMA* 302, 1107–1109. doi: 10.1001/jama.2009.1301

Nature Editorial. (2018). Science needs to redefine excellence. *Nature* 554, 403–404. doi: 10.1038/d41586-018-02183-y

Oaksford, M., and Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Can. J. Exp. Psychol.* 58, 75–85. doi: 10.1037/h0085798

Oaksford, M., and Hahn, U. (2007). "Induction, deduction, and argument strength in human reasoning and argumentation," in *Inductive Reasoning*, eds A. Feeney and E. Heit (Cambridge: Cambridge University Press), 269–301.

Over, D. (2016). "The paradigm shift in the psychology of reasoning," in *Cognitive Unconscious and Human Rationality*, eds L. Macchi, M. Bagassi, and R. Viale (Cambridge, MA: MIT Press), 79–99.

Paulus, F. M., Rademacher, L., Schäfer, T. A. J., Müller-Pinzler, L., and Krach, S. (2015). Journal Impact factor shapes scientists' reward signal in the prospect of publication. *PLoS One* 10:e0142537. doi: 10.1371/journal.pone.0142537

PLoS Medicine Editorial (2006). The impact factor game. *PLoS Med.* 3:e291. doi: 10.1371/journal.pmed.0030291

Reich, E. S. (2013). Science publishing: the golden club. *Nature* 502, 291–293. doi: 10.1038/502291a

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

Simons, K. (2008). *The Misused Impact Factor*. New York, NY: Science.

Sloman, S., and Fernbach, P. (2017). *The Knowledge Illusion: Why We Never Think Alone*. New York, NY: Riverhead Books.

Stephan, P., Veugelers, R., and Wang, J. (2017). Reviewers are blinkered by bibliometrics. *Nature* 544, 411–412. doi: 10.1038/544411a

Tort, A. B. L., Targino, Z. H., and Amaral, O. B. (2012). Rising publication delays inflate journal impact factors. *PLoS One* 7:e53374. doi: 10.1371/journal.pone.0053374

Van Dijk, D., Manor, O., and Carey, L. B. (2014). Publication metrics and success on the academic job market. *Curr. Biol.* 24, R516–R517. doi: 10.1016/j.cub.2014.04.039

Van Noorden, R. (2017). The science that's never been cited. *Nature* 552, 162–164. doi: 10.1038/d41586-017-08404-0

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics* 59, 467–472. doi: 10.1023/B:SCIE.0000018543.82441.f1

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

Werner, R. (2015). The focus on bibliometrics makes papers less useful. *Nature* 517, 245–245. doi: 10.1038/517245a

Wissenschaftsrat [German Council of Science and Humanities] (2015). *Empfehlungen Zu Wissenschaftlicher Integrität*. Stuttgart: Positionspapier.

Yeung, A. W. K. (2017). Do neuroscience journals accept replications? A survey of literature. *Front. Hum. Neurosci.* 11:468. doi: 10.3389/fnhum.2017.00468