Journal of Informetrics 11 (2017) 704–712

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Quantifying perceived impact of scientific publications

Filippo Radicchi*, Alexander Weissman, Johan Bollen

Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, 919 E 10th Street, Bloomington, IN 47408, USA

ARTICLE INFO

Article history:
Received 7 March 2017
Received in revised form 23 May 2017
Accepted 23 May 2017

ABSTRACT

We report on an empirical verification of the degree to which citation numbers represent scientific impact as it is actually perceived by experts in their respective field. We run a survey of about 2000 corresponding authors who performed a pairwise impact assessment task across more than 20,000 scientific articles. Results of the survey show that citation data and perceived impact do not align well, unless one properly accounts for psychological biases that affect the opinions of experts with respect to their own papers vs. those of others. Researchers tend to prefer their own publications to the most cited papers in their field of research. There is only a mild positive correlation between the number of citations of top-cited papers and expert preference in pairwise comparisons. This also applies to pairs of papers with several orders of magnitude differences in their total number of accumulated citations. However, when researchers were asked to choose among pairs of their own papers, thus eliminating the bias favouring one's own papers over those of others, they did systematically prefer the most cited article. We conclude that, when scientists have full information and are making unbiased choices, expert opinion on impact is congruent with citation numbers.

1. Introduction

Metrics based on bibliographic data increasingly inform important decision-making processes in science, such as hiring, tenure, promotion, and funding (Bornmann & Daniel, 2006; Bornmann, Mutz, Marx, Schier, & Daniel, 2011; Bornmann, Wallon, & Ledin, 2008; Harzing, 2010; Hornbostel, Böhmer, Klingsporn, Neufeld, & von Ins, 2009; Lovegrove & Johnson, 2008). Many, if not most, of these bibliometric indicators are derived from article-level citation counts (Bar-Ilan, 2008; Van Noorden, 2010), e.g. the *h*-index (Hirsch, 2005) and the journal impact factor (Garfield, 2006), due to the wide availability of extensive bibliographic records. However, they rest on the frequently undeclared assumption that citations are an accurate and reliable reflection of scientific impact. Bornmann and Daniel (2008) and Sarigöl, Pfitzner, Scholtes, Garas, and Schweitzer (2014) find that citations can serve as reasonable approximations of social indicators such as popularity or success, but they do not support the most important assumption underlying most work in the area of citation-based impact indicators, namely that *citations quantify scientific impact*. Here, we investigate this assumption directly by determining whether citation data truly reflect scientific impact as it is perceived by expert scholars in their respective fields. Such an assumption is central in the long-standing debate about the proper use of citation data (Garfield, 1979). One common argument against the use of citation data in assessment exercises rests in particular on the doubts about their validity as indicators of true scientific

* Corresponding author.
E-mail addresses: filiradi@indiana.edu (F. Radicchi), alexw@indiana.edu (A. Weissman), jbollen@indiana.edu (J. Bollen).

http://dx.doi.org/10.1016/j.joi.2017.05.010
1751-1577/

impact (Adler, Ewing, & Taylor, 2009; MacRoberts & MacRoberts, 1996, 2010). In fact, the literature frequently confounds "impact", "influence", and "rank" with citation-derived metrics (Radicchi, Fortunato, & Castellano, 2008; Wang, Song, & Barabási, 2013; Wuchty, Jones, & Uzzi, 2007) although these metrics can vary along multiple distinct dimensions (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Bornmann & Daniel, 2008).

Determining whether citations actually indicate scientific impact is an empirical question which cannot be resolved by theoretical discussions of the perceived or presumed benefits *vs*. demerits of citation data alone. For this reason, we decided to define and empirically quantify a novel post-publication metric, namely impact as perceived by the authors themselves. Our goal is to understand whether citation numbers "truly" reflect a particular dimension of impact: the influence or importance of papers in the daily practice of researchers.

We designed a large-scale survey at bigscience.soic.indiana.edu to collect responses from thousands of experienced scholars in a multitude of disciplines (Fig. A1). We asked researchers to make a pairwise decision with regards to their preference of one paper over the other. These decisions were made under two conditions: whether the articles were written by the expert scholars themselves or by other scholars. We then aggregated results over the entire population of respondents to quantify the degree of correlation between the pairwise preferences of respondents (i.e., perceived impact) and the actual difference in the number of citations accumulated (i.e., citation impact) for the pair of papers. Our results indicate that, from the perspective of individual researchers, their assessment of impact vs. impact judged from citation data are *only* related for pairs of their own papers. Every time that a paper not co-authored by the individual is involved in the estimation of perceived impact, this comparison shows null or negative correlation with citation impact.

2. Methods

To build the infrastructure needed for the survey, we took all scientific articles with a publication year up to 2013 from the Web of Science (WoS) database. We associated every article with the total number of citations accumulated until 2013 in the WoS citation network as an indication of its citation impact. We identified all articles that were associated with corresponding author(s) from three major public universities in the US: (1) Indiana University (IU), (2) University of Michigan (UMICH), and (3) University of Minnesota (UMN). We chose IU to run the first pilot experiment since it is our home university; this would make initial data validation more straightforward. UMICH and UMN were selected since they have the largest number of recent publications among all public universities in the US. All three universities host departments in almost all disciplines of the natural and social sciences, offering a relatively unbiased sample of participants in terms of field of expertise. Articles were matched to their corresponding author(s) using the email address(es) provided in the article metadata. For example, email addresses ending in "indiana.edu" were taken as an indication that the authors were based at IU. Similarly for UMICH and UMN, we identified researchers from those institutions by retaining email addresses that ended with "umich.edu" and "umn.edu," respectively.

In our data set, articles published prior to 1995 did not provide author email addresses. After 1995, the proportion of articles carrying at least one email address increased rapidly each year. The vast majority of recent articles provide at least one email address. This may introduce a potential source of bias in our study towards relatively recent publications, but this procedure generated a number of very important advantages. First, the association between articles and physical persons was virtually free from homonymy-induced errors. Second, and very important for our purpose, email addresses allowed us to directly contact researchers and invite them to participate in our large-scale survey.

We sent an email message that contained a unique and customized URL to every potential participant (Fig. A2). This URL pointed to a web page in which the respondent was presented with a maximum of ten pairwise comparisons between scientific papers. In each comparison, we provided only the journal and year of publication of the papers, their title and list of authors. For every pair of papers, the respondent was asked to select the one article she/he believed to be more "influential" for her/his own research. Note that this task did not involve any consideration of, nor information on, citation data for any of the two papers. We tailored the survey for every respondent; papers were selected from three different pools that were constructed using information from the publication and citation record of the respondent: (i) "Own" publications (OWN), (ii) "Top cited" articles (TCD), and (iii) "Random" papers (RND). The OWN pool consisted of articles written by the respondents themselves, i.e., they were associated with the email address of the responding author. The pool of TCD articles was constructed as follows. First, we identified all articles appearing at least once in the reference lists of publications by the respondent. We eliminated from this set articles that appeared also in the pool of OWN articles. We then ranked the remaining articles based on their citation impact, and selected the top 10 articles in the list. This procedure allowed us to populate the TCD pool with the most cited articles in the respondent's area of research, but that were not written by the respondent themselves. The pool of RND papers was simply generated as the union of articles (co-)authored or cited (not just the top cited) by all potential participants, thus comprising mainly articles unknown to the respondent. This pool was used only for the IU sample as a control set to check for the presence of possible systematic biases in the survey. Several respondents were confused when presented with random papers. We therefore decided to remove the RND pool from subsequent surveys made at UMICH and UMN. Although the IU respondents reported discomfort about having to make selections from the RND pool of papers, the data generated was useful to validate construction of the other two pools, and to test for the absence of systematic biases in the visual format used in the online survey (Fig. A3). Once the three article pools were generated, every comparison presented to respondents in their personalized survey was composed of pairs of papers taken at random from the various pools. This allowed us to collect information about the preferences of respondents

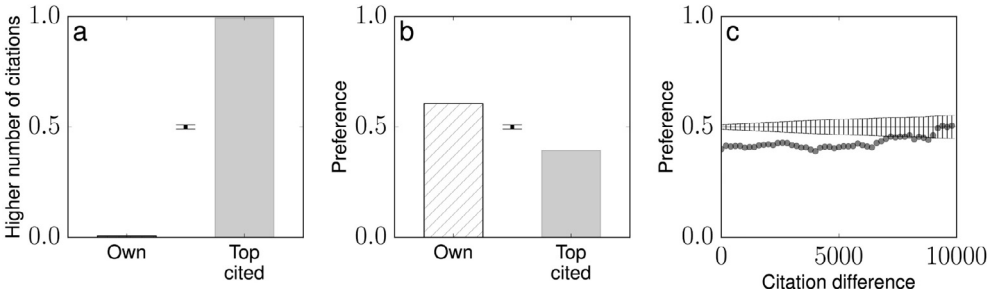

Fig. 1. Perceived impact of authored *vs.* top-cited papers. We consider $N = 2916$ comparisons between one paper taken from the "Own" pool, and the other one extracted from the "Top cited" pool. (a) Probability $P(c_a > c_b | p_a = o, p_b = t)$ that the paper written by the respondent accumulated more citations than the top-cited article. To estimate such a probability, we consider only comparisons between papers with different number of citations, so that the difference in citations is always different from zero. We obtain $P(c_a > c_b | p_a = o, p_b = t) = 0.005$, corresponding to a standard score $|z| = 53.36$ with respect to the unbiased binomial model. The error bar in the graph is centered at the unbiased expected value 0.5, and has height equal to twice the value of the standard deviation $\sigma = 0.009$. (b) Probability $P(a \to b | p_a = o, p_b = t)$ that the respondent preferred her/his own paper to the one taken from the pool of Top cited articles. Such a probability amounts to $P(a \to b | p_a = o, p_b = t) = 0.61$, corresponding to $|z| = 11.48$. (c) Probability $P_{ot}(a \to b | c_a > c_b, c_a - c_b \geq \Delta c)$ that the respondent preferred the paper with more citations. Each point represents the probability of preference for the paper with higher number of citations for all pairs of papers whose citation difference Δc is higher than the value reported on the x-axis. The horizontal line indicates the naive expectation 0.5 of the unbiased binomial model. Height of the error bars equals two standard deviations of the binomial model.

among pairs of papers within the same pool and across different pools. We recorded the preferences expressed by every participant, and we used it to estimate the perceived impact of one paper with respect to the other in the comparison.

We sent emails for participation to a total of 19,546 researchers (Table A1). 1819 scholars participated in the survey, for a total of 12,000 pairwise comparisons among 20,661 distinct articles. This is a low but acceptable response rate for an online survey with a self-selected sample (Sheehan, 2001). Importantly, we did not observe any systematic bias in the pool of participants in terms of academic age (Fig. A4), although we cannot exclude the presence of all selection bias in the sample of researchers who participated in the survey (Bethlehem, 2010). We remark that our estimate of the response rate is conservative; some of the email addresses we used to contact researchers might be no longer active, e.g., through retirement or change of affiliation. Note that the resulting survey sample was more than ten times larger than that of a recent attempt to characterize the features of the top 10 most cited scientific publications authored by biomedical researchers (Ioannidis, Boyack, Small, Sorensen, & Klavans, 2014).

3. Results

3.1. Perceived impact of authored vs. top-cited papers

Fig. 1 summarizes the relation between citation and perceived impact obtained from the analysis of comparisons between one paper taken from the OWN and the other from the TCD pool. Naturally, a paper in the TCD pool is very likely to have a citation impact larger than an article of the OWN, as the measure of the probability $P(c_a > c_b | p_a = o, p_b = t)$ shows in Fig. 1a. Here, c_x denotes the total number of citations accumulated by paper x, and p_x denotes the pool of the paper x (o stands for OWN, and t for TCD). Nonetheless, perceived impact of articles written by respondents themselves is generally higher than the ones of papers from the TCD pool, revealing a strong bias towards respondents' own papers. This is visualized in Fig. 1B, where we consider the probability $P(a \to b | p_a = o, p_b = t)$, i.e., the probability that an OWN article was preferred to one from the TCD pool (the notation $x \to y$ denotes preference of paper x with respect to paper y). To quantify the statistical significance of our measurements, we use absolute values of the standard score z with respect to an unbiased binomial distribution, hence $z = (P - 0.5)/\sigma$, where P is the actual value of the probability measured in the experiment, $\sigma = \sqrt{0.5 \times 0.5/N}$, and N is the size of the sample (i.e., number of comparisons). Note that, since only two possibilities are available, it does not matter if we measure the probability P or its complementary probability $1 - P$. The absolute value of the standard score $|z|$ does not depend on this choice. The statistical interpretation of the standard score for an unbiased binomial distribution with sufficiently large sample sizes, as in our case, is similar to the one valid for the standard normal distribution, so that p-values <0.001 approximately correspond to $|z| > 3$, and the p-value decreases exponentially fast as $|z|$ increases. The empirical results of Fig. 1B are highly unlikely to happen by chance. In fact, we have $P(a \to b | p_a = o, p_b = t) = 0.61$ and $N = 2916$, leading to $|z| = 11.48$. One may wonder whether this result is actually dependent on the difference in citation impact between the two papers or not. Fig. 1C points to a possible answer to this question. We consider the probability $P_{ot}(a \to b | c_a > c_b, c_a - c_b \geq \Delta c)$ that respondents preferred the more cited paper among the two articles in the comparison as a function of the difference Δc in citations between the two papers (for brevity we used the suffix ot to indicate the pools where the two papers were taken from). Surprisingly, citation and perceived impact are negatively correlated, in a statistically significant manner, for a wide range of values of the difference of citations accumulated by the two papers. Only when the citation impact of the TCD paper is much larger than the one of the OWN paper, we do not longer observe a statistically significant preference. The fact that scholars tended to systematically prefer their own articles regardless of the comparison can be interpreted as the

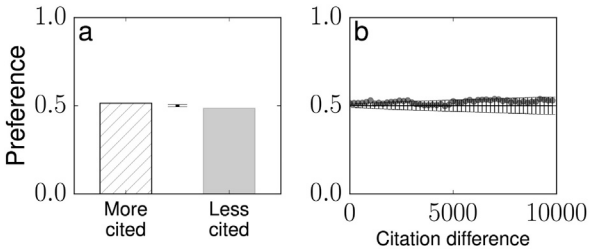

Fig. 2. Perceived impact of top-cited papers. We consider only comparisons between two papers taken from the "Top cited" pool. (a) Probability $P_{tt}(a \to b|c_a > c_b)$ that the respondent preferred the paper with higher/lower number of citations. Papers with higher citation counts are preferred with probability $P_{tt}(a \to b|c_a > c_b) = 0.51$, calculated over a sample of size $N = 5930$, leading to a standard score $|z| = 2.18$. Error bar is centered at 0.5 and has height equal to twice the standard deviation σ of the unbiased binomial model. Here $\sigma = 0.007$. (b) Probability $P_{tt}(a \to b|c_a > c_b, c_a - c_b \geq \Delta c)$ that the respondent preferred the paper with more citations as a function of the difference in citations among the papers. The horizontal line indicates the naive expectation of 0.5 of the unbiased binomial model. Height of the error bars equals two standard deviations of the binomial model.

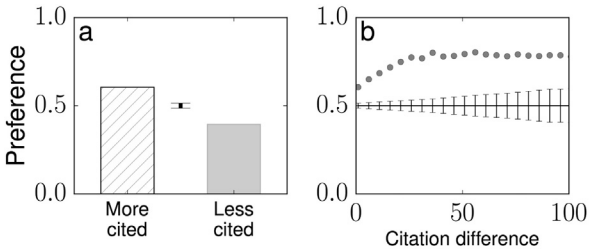

Fig. 3. Perceived impact of authored papers. We consider only comparisons between two papers taken from the "Own" pool. (A) Probability $P_{oo}(a \to b|c_a > c_b)$ that the respondent preferred the paper with higher or lower number of citations. Papers with higher citation counts are preferred with probability $P_{oo}(a \to b|c_a > c_b) = 0.61$ calculated over a sample of size $N = 1302$, leading to a standard score $|z| = 7.60$. The error bar is centered at 0.5 and has a height equal to twice the standard deviation σ of the unbiased binomial model. Here $\sigma = 0.014$. (B) Probability $P_{oo}(a \to b|c_a > c_b, c_a - c_b \geq \Delta c)$ that the respondent preferred the paper with more citations as a function of the difference in citations among the papers. The horizontal line indicates the naive expectation of 0.5 of the unbiased binomial model. Height of the error bars equals two standard deviations of the binomial model.

consequence of an egocentric (Ross & Sicoly, 1979) or familiarity bias (Zajonc, 1968). Egocentric bias does not necessarily have a negative connotation. The bias could be reconcilable with the fact that researchers base most of their work on results of their own past research, and they might have interpreted the generic question posed in the survey in this way. Furthermore, since researchers are inherently most familiar with their own work, the uncertainty with regards to their impact might be lower and hence these articles might enjoy the authors' preference over a paper that is less familiar and whose impact is thus more difficult to assess reliably.

3.2. Perceived impact of top-cited papers

To discount for the presence of an egocentric or familiarity bias, we turn our attention to results obtained from comparisons where both papers were extracted from the same pool. We start from the description of our findings regarding the TCD pool (Fig. 2). Here, a positive correlation is observed between citation impact and perceived impact. Overall, papers with higher citation impact were preferred at higher rates by researchers, but only with probability $P_{tt}(a \to b|c_a > c_b) = 0.51$ leading to a standard score $|z| = 2.18$ (p-value = 0.02, Fig. 2a). Also, perceived impact does not vary significantly with citation impact even if the difference in citations received by papers in the comparison can be larger than three orders of magnitude (Fig. 2b). The result seems not dependent on the age of papers, as we do not observe any systematic preference (Fig. A5). We speculate that, from the perspective of individual researchers, citing a top-cited article in their own research area may be interpreted as a public "homage" to the collective popularity of the paper, and thus reconcilable with a cumulative advantage principle (Barabási & Albert, 1999; Price, 1976; Wang et al., 2013), the conformity bias (Asch, 1956), or a copying behaviour (Krapivsky & Redner, 2005; Simkin & Roychowdhury, 2002), rather than a true recognition of the importance of the work for their own research.

3.3. Perceived impact of authored papers

Figs. 1 and 2 suggest that the total number of citations received by a paper does not reliably quantify the true impact or influence that the paper has for the actual research of a scholar. Such a conclusion is, however, overturned when we consider pairwise comparisons between articles from the OWN pool (Fig. 3). In this case, we observe a significant alignment between citation impact and expert impact as researchers systematically preferred the highest cited articles among their own articles. Preference to the more cited article was given with probability $P_{oo}(a \to b|c_a > c_b) = 0.61$, corresponding to $|z| = 7.60$ (Fig. 3a).

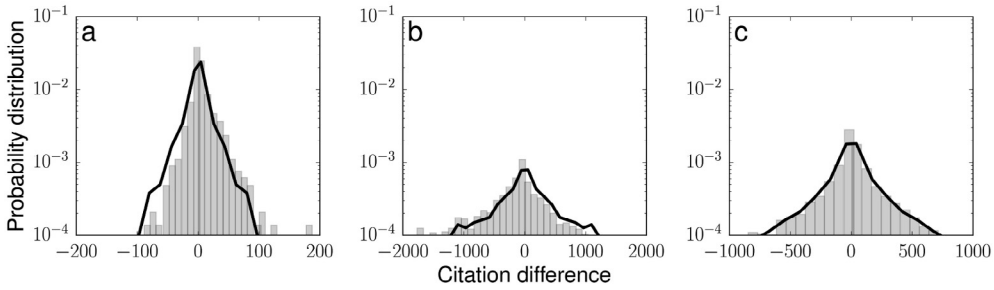

Fig. 4. Citation impact of articles with higher perceived impact. (a) We consider only comparisons between two papers taken from the "Own" pool as in Fig. 3. We include here also comparisons between papers having the same number of citations, so that the difference of their citation impacts can be equal to zero. The gray bars stand for the probability distribution $P_{oo}(c_a - c_b = \Delta c | a \to b)$ of the difference in citation impact between preferred and non-preferred article in the comparison. The average value of the distribution is 11.16, the standard deviation is 130.50, and the skewness is 29.28. The black line serves as a term of comparison as it represents the distribution $P_{oo}(c_a - c_b = \Delta c)$ of difference in citation impact between pairs of papers in the comparisons, irrespective of the preference expressed by researchers. This distribution is by definition symmetric (null skewness), and centered in zero (average value equal to zero). Probability distributions are normalized such that the integral below the curves equals one. (b) Same as in panel a but for comparisons between one paper taken from the "Own" pool and the other extracted from the "Top cited" pool. Data are the same as those used in Fig. 1. The average value of the distribution $P_{ot}(c_a - c_b = \Delta c | a \to b)$ is −267.50, the standard deviation is 4762.69, and the skewness is −1.36. The black line stands for the unconditional probability $P_{ot}(c_a - c_b = \Delta c)$. (c) Same as in panels A and B but for comparisons between two papers taken from the "Top cited" pool. Data are the same as those used in Fig. 2. The average value of the distribution $P_{tt}(c_a - c_b = \Delta c | a \to b)$ is 99.47, the standard deviation is 4255.18, and the skewness is 12.34. The black line stands for the unconditional probability $P_{tt}(c_a - c_b = \Delta c)$.

Furthermore, we observed a systematic increase in the preference for the more cited paper as a function of the difference of citation impact between the papers in the comparison (Fig. 3b). Preference values saturate to almost 0.75, if citations accumulated by papers differ by 50 or more. This effect is even stronger when one accounts for the general tendency of preference for more recent publications, and the fact that more recent publications generally exhibit lower citation impact (Fig. A6). Overall, this finding conclusively supports the observations by Ioannidis et al. in their small-scale survey (Ioannidis et al., 2014).

3.4. Citation impact of articles with higher perceived impact

Our observations are further supported by the results presented in Fig. 4, where we consider the distribution of the difference in citation impact between preferred and non-preferred articles in different types of comparisons. The distribution is clearly right-skewed for comparisons between papers in the pool of OWN articles (Fig. 4a). The distribution for comparisons between articles from the OWN and TCD pools on the other hand has negative average value, and is negatively skewed (Fig. 4b). Finally, the distribution in Fig. 4c, for comparisons between papers in the TCD pool, provides evidence of a slight preference for articles with higher number of citations, but the preference is much less evident than in the case of Fig. 4a.

4. Conclusions

In summary, our work quantifies the degree of correlation between citation impact and a new post-publication metric, namely impact as perceived by the authors themselves. Our survey serves to understand whether citation numbers "truly" reflect the impact of papers as it pertains to the daily practice of researchers. This is a particular and important dimension of impact that has been not quantified before.

In the present study, we neglected potential confounding factors that may affect the perception of importance of a paper over another. These include article's factors as year and journal of publication. Also factors arising from the authorship of papers may play an important role. For example, the preference of a respondent may be biased towards a paper written by a collaborator or towards authors affiliated with prestigious institutions. Additional experiments are needed to properly quantify the effect of these potentially biasing factors.

Provided that the above-mentioned factors do not significantly alter the outcomes of the experiments we conducted in the present study, the results from our large-scale survey generate one very important conclusion: citation numbers approximate with good accuracy the perceived impact of scientific publications. This conclusion can be justified on the basis our results on comparisons among pairs of one's own papers, and the reasonable assumption that the self-assessment of one's own papers is the most objective (and least biased) quantification of perceived impact since authors are assumed to know their own articles best and are thus the best judges of their comparative impact. Another possible explanation could be the Goodhart effect (Goodhart, 1984), namely that the preferences of authors are shaped by the knowledge of citations accumulated by their own papers. As a result, pairwise assessments of their own papers will be congruent with citation data.

Author contributions

Conceived and designed the analysis: Filippo Radicchi, Alexander Weissman and Johan Bollen.
Collected the data: Filippo Radicchi, Alexander Weissman.
Contributed data or analysis tools: Filippo Radicchi, Alexander Weissman.
Performed the analysis: Filippo Radicchi.
Wrote the paper: Filippo Radicchi and Johan Bollen.

Acknowledgements

We are grateful to all researchers who took part in our survey. This work uses Web of Science data by Thomson Reuters, provided by the Network Science Institute at Indiana University. This work is funded by the National Science Foundation (grant SMA-1446078 and SMA-1636636).

Appendix A

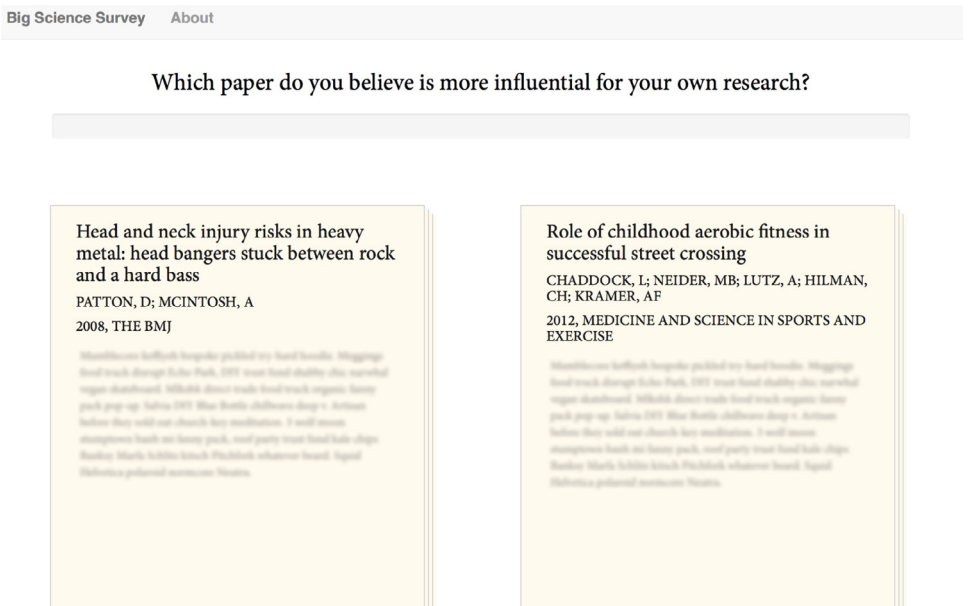

Fig. A1. Survey design. Screenshot from the online survey at bigscience.soic.indiana.edu. The visualization shows an example of a comparison presented to individual participants. Papers visualized in this example were randomly chosen from our dataset.

Table A1
Participation in the survey. Summary table listing number of potential participants, number of respondents, and response rate for individual institutions. In the last row of the table we list values of the same quantities for the entire survey.

Institution	Potential participants	Respondents	Response rate
Indiana University	2673	313	11.7%
University of Michigan	9560	889	9.3%
University of Minnesota	7313	617	8.4%
	19,546	1840	9.4%

```
Hello,

You are receiving this email because you are the
corresponding author of several scientific publications.
We are conducting a survey, the largest of its kind,
to determine the relative importance of scientific
publications as perceived by experienced researchers.

We've generated a survey specifically for you that
consists of anywhere between 1 and 10 pairs of papers
drawn from our database. The publications may be drawn
from your own publications or publications you've cited.
For each pair, we ask you to select the publication that
you believe is more influential for your own research.

It should take less than two minutes to complete the
ten comparisons. To participate, please follow this link:

bigscience.soic.indiana.edu/access-token/{{access_token}}

This survey is a step towards critically assessing the
way the scientific community measures the impact of
its work.  We hope you will participate and help further
this important research.

This survey is being conducted by Alexander Weissman and
Filippo Radicchi from the School of Informatics at
Indiana University. The research is sponsored by
the NSF (Award #1446078) and approved by the Indiana
University IRB (Protocol #1407480218). For more information,
please write to Filippo Radicchi: filiradi@indiana.edu.

With regards,

Alexander Weissman and Filippo Radicchi
School of Informatics and Computing
Indiana University
```

Fig. A2. Email message for participation in the survey. We sent this message to all potential participants in our survey. The email message contained a unique token key for every individual participant, pointing to her/his customized survey.

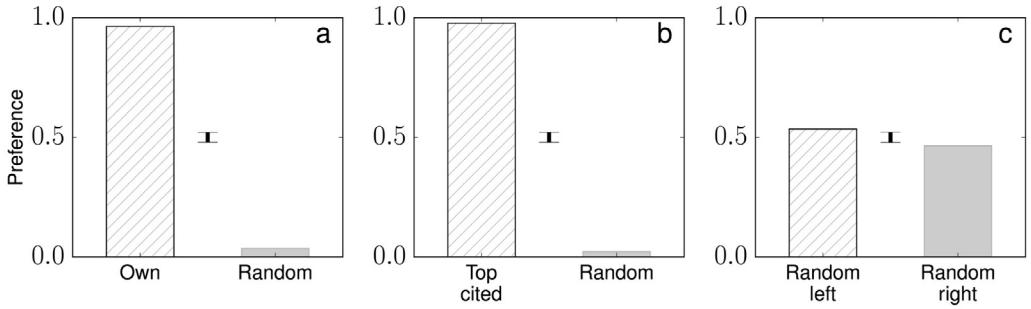

Fig. A3. Sensitivity analysis. Results of this figure are based on the experiment we conducted at Indiana University, the only experiment where papers from the Random pool were considered. (a) Probability $P(a \to b|p_a = o,\ p_b = r)$ that the respondent preferred the paper from the Own pool instead of the paper from the Random pool. Such a probability is $P(a \to b|p_a = o,\ p_b = r) = 0.97$ measured over $N = 587$ total comparisons, and leading to a standard score $|z| = 22.50$. (b) Same as in panel A but for comparisons between one paper from the Top cited pool and the other from the Random pool. In this case, we have: $P(a \to b|p_a = t,\ p_b = r) = 0.98$, $N = 578$, and $|z| = 22.96$. (c) Results for comparisons where both papers were taken from the Random pool. We measure the probability $P(a \to b|g_a = L,\ g_b = R)$ that the respondent selected the paper appearing on the left ($g_x = L$) or right ($g_x = R$) side of the screen during the survey. We have: $P = 0.53$, $N = 550$, and $|z| = 1.62$.

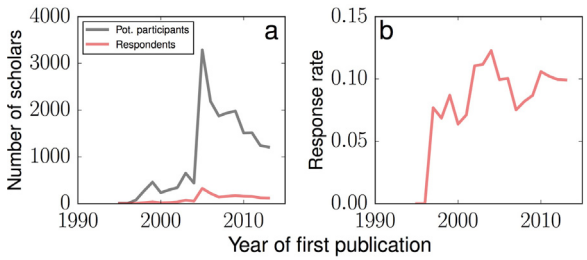

Fig. A4. Participation in the survey. (a) Total number of potential participants (gray line) and respondents (red line) to the survey as functions of the year of their first publication appearing in our database. The first publication corresponds to the oldest paper where the email address of the scholar was appearing in the article metadata. (b) Response rate (i.e., number of respondents divided by number of potential participants) as a function of the year of their first publication in the database.

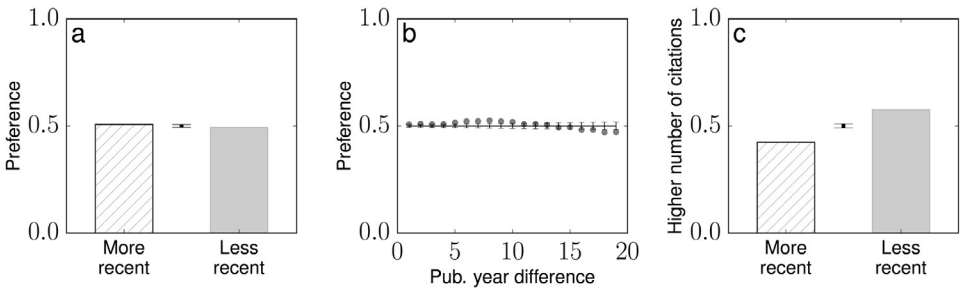

Fig. A5. Perceived impact of top-cited publications depending on the age of papers. We consider only comparisons between two papers taken from the "Top-cited" pool as in Fig. 2 of the main text. (a) Probability $P_{tt}(a \to b|y_a > y_b)$ that the respondent preferred the more/less recent paper (y_x is the year of publication of paper x). To estimate the probability, we consider only comparisons between papers with different years of publication, so that their difference is different from zero. More recent papers are preferred with probability $P_{tt}(a \to b|y_a > y_b) = 0.51$ calculated over a sample of size $N = 5609$, leading to a standard score $|z| = 1.05$. Error bar is centered at 0.5 and has height equal to twice the standard deviation σ of the unbiased binomial model. Here $\sigma = 0.007$. (b) Probability $P_{tt}(a \to b|y_a > y_b, y_a - y_b \geq \Delta y)$ that the respondent preferred the more recent paper among the two in the comparison as a function of the difference of year of publication of the two papers. The horizontal line indicates the naive expectation of 0.5 of the unbiased binomial model. Error bars stand for standard deviation of the binomial model. (c) Probability $P_{tt}(c_a > c_b|y_a > y_b)$ that the more/less recent paper in the comparison has accumulated more citations. To estimate the probability, we considered only comparisons between papers with different years of publication and different citation numbers, so that the difference of both these numbers is different from zero. The probability is $P_{tt}(c_a > c_b|y_a > y_b) = 0.42$, calculated over a sample of size $N = 2742$, leading to a standard score $|z| = 8.02$. Standard deviation computed according to the unbiased binomial model is $\sigma = 0.009$.

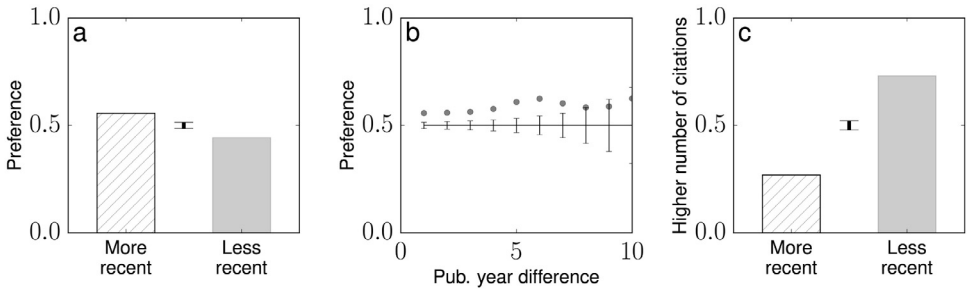

Fig. A6. Perceived impact of authored publications depending on the age of papers. We consider only comparisons between two papers taken from the "Own" pool as in Fig. 3 of the main text. (a) Probability $P_{oo}(a \to b | y_a > y_b)$ that the respondent preferred the more/less recent paper (y_x is the year of publication of paper x). To estimate the probability, we consider only comparisons between papers with different years of publication, so that their difference is different from zero. More recent papers are preferred with probability $P_{oo}(a \to b | y_a > y_b) = 0.56$ calculated over a sample of size $N = 1207$, leading to a standard score $|z| = 3.94$. Error bar is centered at 0.5 and has height equal to twice the standard deviation σ of the unbiased binomial model. Here $\sigma = 0.014$. (b) Probability $P_{oo}(a \to b | y_a > y_b, y_a - y_b \geq \Delta y)$ that the respondent preferred the more recent paper among the two in the comparison as a function of the difference of year of publication of the two papers. The horizontal line indicates the naive expectation of 0.5 of the unbiased binomial model. Error bars stand for standard deviation of the binomial model. (c) Probability $P_{oo}(c_a > c_b | y_a > y_b)$ that the more/less recent paper in the comparison has accumulated more citations. To estimate the probability, we considered only comparisons between papers with different years of publication and different citation numbers, so that the difference of both these numbers is different from zero. The probability is $P_{oo}(c_a > c_b | y_a > y_b) = 0.27$, calculated over a sample of size $N = 550$, leading to a standard score $|z| = 10.83$. Standard deviation computed according to the unbiased binomial model is $\sigma = 0.021$.

References

Adler, R., Ewing, J., & Taylor, P. (2009). Citation statistics. *Statistical Science*, *24*(1), 1–14, 02

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, *70*, 1–70.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century – A review? *Journal of Informetrics*, *2*(1), 1–52.

Bethlehem, J. (2010). Selection bias in web surveys? *International Statistical Review*, *78*(2), 161–188.

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, *4*(6), e6022.

Bornmann, L., & Daniel, H. D. (2006). Selecting scientific excellence through committee peer review – a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants? *Scientometrics*, *68*(3), 427–440.

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.

Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H. D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 857–879.

Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. *PLoS ONE*, *3*(10), 3480.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375.

Garfield, E. (2006). The history and meaning of the journal impact factor? *Journal of the American Medical Association*, *295*(1), 90–93.

Goodhart, C. A. (1984). *Problems of monetary management: The UK experience*. Springer.

Harzing, A. W. (2010). *The publish or perish book*. Tarma Software Research.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 6569.

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence? *Scientometrics*, *79*(1), 171–190.

Ioannidis, J., Boyack, K. W., Small, H., Sorensen, A. A., & Klavans, R. (2014). Is your most cited work your best? *Nature*, *514*(7524), 561–562.

Krapivsky, P., & Redner, S. (2005). Network growth by copying. *Physical Review E*, *71*(3), 36118.

Lovegrove, B. G., & Johnson, S. D. (2008). Assessment of research performance in biology: How well do peer review and bibliometry correlate? *Bioscience*, *58*(2), 160–164.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis? *Scientometrics*, *36*(3), 435–444.

MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences? *Journal of the American Society for Information Science and Technology*, *61*(1), 1–12.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes? *Journal of the American society for Information science*, *27*(5), 292–306.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact? *Proceedings of the National Academy of Sciences United States America*, *105*(45), 17268–17272.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*(3), 22.

Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). *Predicting scientific success based on coauthorship networks*. arXiv:1402.7268

Sheehan, K. B. (2001). E-mail survey response rates: A review? *Journal of Computer-Mediated Communication*, *6*(2), 0–0.

Simkin, M. V., & Roychowdhury, V. P. (2002). *Read before you cite!*. https://arxiv.org/abs/cond-mat/0212043

Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature*, *465*(7300), 864–866.

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2p2), 1.