Research Policy 51 (2022) 104484

Contents lists available at ScienceDirect

Research Policy

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

How status of research papers affects the way they are read and cited

Misha Teplitskiy[a,e,*], Eamon Duede[b,c,g], Michael Menietti[d,e], Karim R. Lakhani[d,e,f]

[a] University of Michigan School of Information, Ann Arbor, MI, 48109, USA
[b] Dept. of Philosophy, University of Chicago, Chicago, IL, 60637, USA
[c] Committee on Conceptual and Historical Studies of Science, University of Chicago, Chicago, IL, 60637
[d] Harvard Business School, Harvard University, Boston, MA, 02163, USA
[e] Laboratory for Innovation Science at Harvard, Harvard University, Boston, MA, 02163, USA
[f] Institute for Quantitative Social Science, Harvard University, Cambridge, MA, 02138, USA
[g] Knowledge Lab, University of Chicago, Chicago, IL, 60637, USA

ARTICLE INFO

JEL codes:
O33
O34

Keywords:
Metrics
Citations
Influence
Science
Status

ABSTRACT

Although citations are widely used to measure the influence of scientific works, research shows that many citations serve rhetorical functions and reflect little-to-no influence on the citing authors. If highly cited papers disproportionately attract rhetorical citations then their citation counts may reflect rhetorical usefulness more than influence. Alternatively, researchers may perceive highly cited papers to be of higher quality and invest more effort into reading them, leading to disproportionately substantive citations. We test these arguments using data on 17,154 randomly sampled citations collected via surveys from 9,380 corresponding authors in 15 fields. We find that most citations (54%) had little-to-no influence on the citing authors. However, citations to the most highly cited papers were 2–3 times more likely to denote substantial influence. Experimental and correlational data show a key mechanism: displaying low citation counts lowers perceptions of a paper's quality, and papers with poor perceived quality are read more superficially. The results suggest that higher citation counts lead to more meaningful engagement from readers and, consequently, the most highly cited papers influence the research frontier much more than their raw citation counts imply.

1. Introduction

A principal means by which scientists acknowledge their intellectual debts to prior research is through the practice of citation (Merton 1988; Zuckerman 1987). In turn, administrators and researchers often use citations, or metrics derived from them like journal impact factors and *h*-indices, to quantify intellectual influence and allocate funding, awards, and promotions (Abbott et al., 2010; McKiernan et al., 2019; Langfeldt et al., 2021). An assumption that is often implicit in these use cases is that citations are similar to one another in the amount of influence they represent.[1] Accordingly, the total influence of a paper on other researchers is usually equated with its overall citation count.

Yet scholars have long argued that citations come in different types (Tahamtan and Bornmann 2019; Cronin 1984; Liu 1993). Undoubtedly, authors cite some works to acknowledge significant intellectual influence on their projects (Baldi 1998; Zuckerman 1987). Nevertheless, authors also cite works for various other reasons, many of which have little to do with acknowledging influence (Cozzens 1989; Gilbert 1977; MacRoberts and MacRoberts 1989). For example, authors may cite works to provide context, to distinguish their contributions from the prior literature, or to criticize that literature (Brooks 1986; Catalini et al., 2015; Liu 1993; Tahamtan and Bornmann 2018). Some citations may even be "coerced" during the publishing process (Wilhite and Fong 2012). According to this literature, citations vary widely in how much influence they represent, which we call their "influence intensity." Consequently, to properly quantify a paper's influence, it is important not only to know how many times it has been cited but also the influence intensity of those citations. For example, among two papers in the same research area, one with 1000 citations may be only slightly more influential than one with 100 if the former attracted primarily citations with low influence intensity.

Therefore, understanding heterogeneity in citing can improve our

* Corresponding author.
E-mail addresses: tepl@umich.edu (M. Teplitskiy), eduede@uchicago.edu (E. Duede), mmenietti@hbs.edu (M. Menietti), klakhani@hbs.edu (K.R. Lakhani).

[1] Alternatively, it may be assumed that citations do vary in how much influence they represent but that variation is more-or-less random. The practical implication of either assumption is to ignore the variation.

https://doi.org/10.1016/j.respol.2022.104484
Received 14 August 2021; Received in revised form 11 January 2022; Accepted 12 January 2022
Available online 18 January 2022

understanding and measurement of scholarly influence. We conduct a large-scale study of how citations vary in influence intensity, using a personalized survey with an embedded experiment responded to by 9380 corresponding authors in 15 academic fields. With these data, we address the following three research questions.

First, *what is the overall variation in influence intensity of citations in the research literature?* In other words, what fraction of citations are "substantive" (high influence intensity) vs. "rhetorical" (low influence intensity).[2] A number of studies investigating citation types have argued that many if not most citations are rhetorical (Horbach et al., 2021; Krampen et al., 2007; MacRoberts and MacRoberts 1989; MacRoberts and MacRoberts 1996). However, these studies typically rely on third parties to label citation types. Without direct access to the underlying motivations behind specific citation decisions, third parties look for evidence of motivations in the written text, usually citations' surrounding context (Tahamtan and Bornmann 2019). If substantive engagement in the written context is not explicit and readily perceived, the citation is assumed to be rhetorical in nature. Consequently, if some references substantively influence authors but they do not make it readily apparent in the text, this research design risks overestimating the frequency of rhetorical citing. Lastly, many studies in this literature typically focus on just one academic field. We contribute to this literature with large-scale, cross-field, and easily interpretable characterization of citation types provided by the individuals most knowledgeable about the motivations behind each citation – the citing authors themselves.

Second, *do highly and lightly cited papers tend to attract citations of different influence intensities?* In other words, are famous papers cited more or less substantively than obscure ones? The existing literature related to this question is smaller and less conclusive. Case studies tend to focus on specific highly cited papers, finding that many of their citations are rhetorical in nature. For example, Star has argued that citations to her classic paper on "boundary objects" tend to miss key characteristics of the concept (Leigh Star 2010). Mizruchi and Fein find that authors cite the classic work by DiMaggio and Powell on institutional isomorphism in ways that are only loosely connected to its contents (Mizruchi and Fein 1999). These and other examples (Simmons et al., 2018) raise the possibility that the lofty citation counts of highly cited papers may reflect those papers' rhetorical usefulness to citing authors more than substantive influence upon them. More systematic studies have been suggestive but indirect, lacking direct measures of influence intensity, and reaching divergent conclusions (Moed and Garfield 2004; Frandsen and Nicolaisen 2017). We contribute to this literature with data and analysis obtained by directly measuring the influence intensity of randomly sampled citations to lightly and highly cited papers.

Lastly, we ask *why highly and lightly cited papers tend to attract citations of different influence intensity?* What are the mechanisms connecting a paper's "citedness" at a particular time to how it is subsequently cited? The existing literature on this question is the least developed. Prior work has focused on how the characteristics of papers and their citedness affect future citation amount (Wang et al., 2013), but not the influence intensity of those citations. We greatly extend this literature by drawing on endogenous crowds theory (Le Mens et al. 2018) and citing theories (Nicolaisen 2007; Bornmann and Daniel 2008) to develop competing accounts for how the citedness of a paper affects the type of its subsequent citations. Using an experiment embedded in the survey and correlational evidence, we show that the causal evidence favors one of the accounts.

Answering these three questions is critical for research management and understanding inequalities in science. For research management, it is important to accurately quantify the influence of specific papers and portfolios of papers. For example, does one 1000-citation paper tend to represent more or less scholarly influence than a portfolio of 10 papers with 100 citations each? Relatedly, better quantification of influence can help elucidate inequality in science. Given the association of citations with influence, inequality in influence is often operationalized as inequality in citations. Studies taking this approach find that the amount of inequality is large and growing (Nielsen and Andersen 2021**)**. However, if the influence and citation equivalence breaks down among prominent or obscure papers, inequality in citations may over- or understate the true inequality in influence. Similar questions apply when citations are used to measure the contribution to scientific progress of a few elite, highly cited researchers vs. the vast majority of non-elite ones (Cole and Cole 1972; Macroberts and Macroberts 1987).

Furthermore, as Robert Merton famously argued, high-status works and actors tend to receive more recognition than low-status ones for products of similar quality (Merton 1968). This pattern is often referred to as the Matthew Effect and cumulative advantage (Allison et al., 1982; Rigney 2010). Empirical evidence broadly supports this pattern in research funding (Bol et al., 2018), manuscript peer review (Sun, Barry Danfa, and Teplitskiy n.d.; Tomkins et al., 2017), and citation amount (Azoulay et al., 2013). However, whether cumulative advantage also occurs through citation type (*i.e.*, whether high-status works attract more or less influential citations) is unclear.

The remainder of this paper is organized as follows. In the next section, we develop competing accounts of how citation intensity varies across highly and lightly cited papers. We then present the data and methods used to collect detailed data on systematically sampled citations. Next, we discuss the overall heterogeneity of citations in terms of their influence intensity, and establish the mechanisms by which citation amount influences the type of subsequent citations. Lastly, we conclude with policy implications and directions for future work. A document with supplementary information (SI) contains additional detail on survey materials and responses, and robustness analyses.

2. Relationship between citation amount and citation type

To develop the possible relationships between citation amount and type, we conceptualize a set of papers as belonging to some research area where the papers vary in characteristics that are relatively stable, like topic, quality, or the journal in which they are published, and papers that vary in characteristics that are time-varying such as citation count. From this set, researchers strategically choose what to read and cite (Renear and Palmer 2009; Rubin and Rubin 2021; Seeber et al., 2019). Drawing on both the normative and social constructivist theories of citing (Nicolaisen 2007; Bornmann and Daniel 2008), we assume the researchers vary in their motivations in making these choices. At any given moment, some researchers are "substantively motivated" and are seeking papers from which to get ideas that will influence their ongoing or future work, ultimately resulting in substantive citations. Others are "rhetorically motivated" and are seeking papers to cite for various rhetorical purposes, such as to support some specific argument in their own work, ultimately resulting in making citations for largely rhetorical reasons. Researchers may have different motivations for different papers or arguments within papers. Understanding how different papers are cited comes down to understanding how rhetorically and substantively motivated researchers allocate their attention across these papers.

We argue that both stable and time-varying characteristics of papers may affect the amount and composition of researcher attention and citations. Briefly, papers with high quality may attract a large number of substantively motivated researchers. After some time, high quality

[2] The substantive-vs-rhetorical distinction departs from the existing literature on citation types (Liu 1993; Tahamtan and Bornmann 2019) in that it captures only one aspect of citations: how much they influenced the citing authors. While more nuanced citation types, including the ones mentioned above, may all serve useful purposes (except for perhaps coerced citations), taken at a sufficiently general level, all the types reflect certain influence intensities and fall somewhere on the substantive-rhetorical spectrum. Importantly, this distinction does not imply that papers cited rhetorically are of lower quality.

papers would have high citation counts and their citations would be on average substantive. In this scenario, the resulting positive association between citation count and substantive citations would not be causal. Citation counts may also play a causal role, although the direction of the effect is ambiguous. On the one hand, as a paper's citation count increases, it may make a paper increasingly attractive to rhetorically motivated researchers, perhaps from neighboring fields. After some time, higher citation counts would be associated with less substantive citations. On the other hand, as a paper's citation count increases, substantively motivated researchers may perceive it to be of higher quality and worth their effort, and, after some time, higher citation counts would be associated with more substantive citations. This citation counts-based process produces a similar empirical prediction to the one based on quality.

Empirically, we focus on whether the most recently received ("marginal") citation is substantive or rhetorical. Focusing on a paper's most recent citation rather than the composition of all citations to date is chosen to best match our survey design, which asks authors about relatively recent citation decisions to minimize errors in recollection. However, the arguments should apply across papers' citation lifetimes, and, empirically, the survey samples older and newer papers, which provide data from different points in papers' lifetimes.

Fig. 1 visualizes the implications of the potential citation amount-type relationships. The *x*-axis in all four panels denotes papers' total accumulated citations, with 100 denoting a hypothetical maximum. Panels A, B, C correspond to cases where the marginal citation is more, less, or equally substantive for more highly cited papers, with the *y*-axis denoting the probability that the marginal citation is substantive. Finally, panel D shows how citation counts may be biased measures of influence: for each of the three cases, it measures a paper's influence as the number of substantive citations only, and compares that to the number of citations of all types.

2.1. Stable characteristics: quality

First, a stable characteristic like quality may be valued by both substantively and rhetorically motivated researchers, but it is likely to be valued more by the former. Researchers read relatively few papers from the set of potentially relevant ones (Tenopir et al., 2015), so they are likely to invest that effort selectively into the best works (Wang and Soergel 1998). A paper's quality is also likely correlated with its overall citation potential. Consequently, higher quality papers may receive disproportionately substantive citations *consistently*, early on and later in their lifetimes. After a certain time, if papers are ranked according to citations accrued, papers with higher citation counts would also have disproportionately substantive citations, as pictured in Fig. 1A. Panel D shows that if this pattern dominates (red curve), high citation counts understate a paper's actual influence.

2.2. Time-varying characteristics: citation count

However, endogenous crowd theory (Le Mens et al. 2018) suggests that the time-varying characteristics of papers, such as their citation counts, may causally affect the amount and composition of subsequent

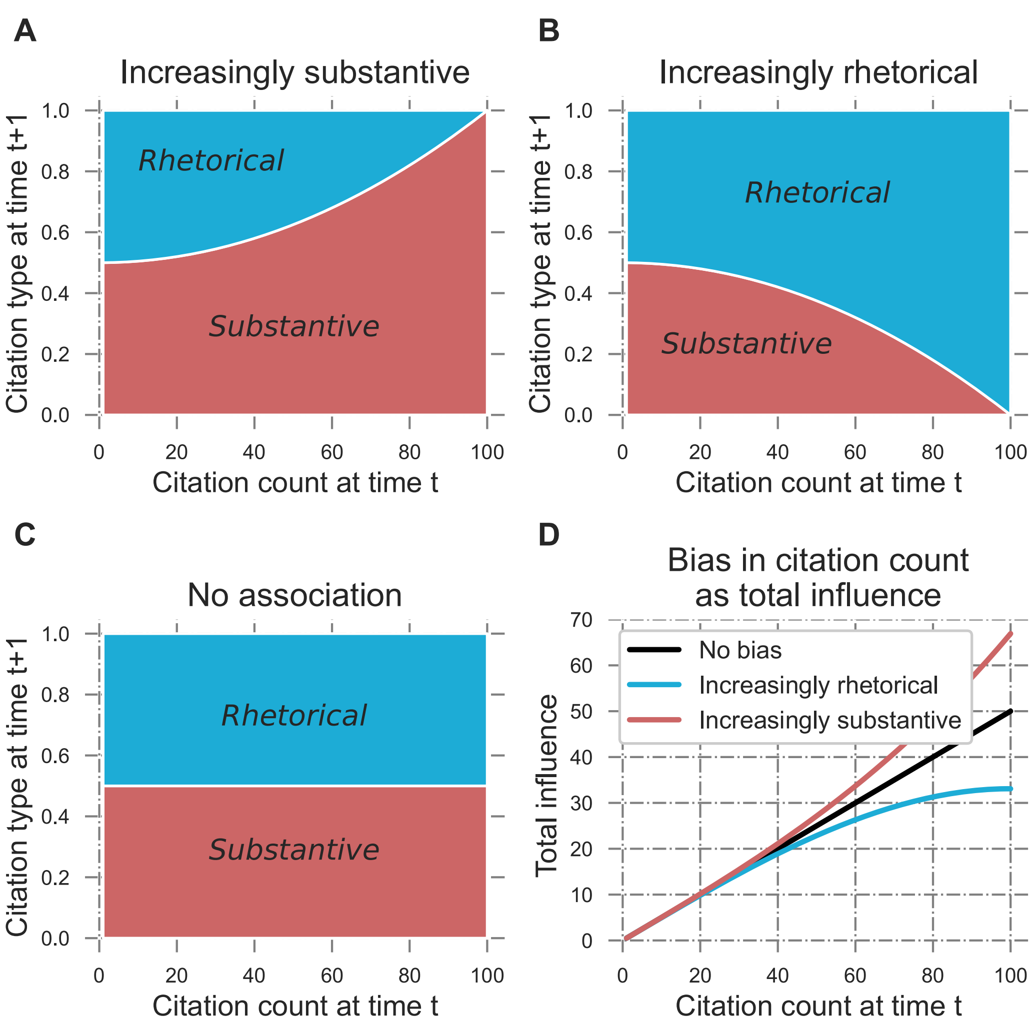

Fig. 1. Three possibilities of how a paper's citation count at time t is related to whether its next citation, at time $t + 1$, is rhetorical or substantive. The x-axis denotes citations accumulated by a large sample of papers published at the same time, with 100 denoting the hypothetical maximum. **A:** The higher the citation count, the more likely the next citation to be substantive. **B**: The higher the citation count, the more likely the next citation is to be substantive. **C**: Citation count and marginal citation type are unrelated. **D**: Implications of these possibilities for how a paper's overall number of citations is related to its actual influence (number of substantive citations).

attention. In our setting, an increasing citation count may have opposite effects. First, researchers may use citation counts as a signal of quality (van Dalen and Henkens 2005; Wang and Soergel 1998). As a paper's citation count grows, the improved perception of quality would make the paper sufficiently attractive to an increasing number of substantively motivated potential citers. There are two mechanisms that would help positively perceived papers achieve influence among the substantively motivated researchers: early enough discovery and close enough reading. First, for a paper to have influence on their research choices, the researchers must discover it early enough in their projects. Positively perceived papers are likely to be more widely presented, taught, and otherwise diffused, making it more likely that the papers are discovered early (Johnson and Oppenheim 2007; Milard 2014; Milard and Tanguy 2018; Murray and Poolman 1982). Second, researchers must read the paper carefully enough to understand its content. Researchers are more likely to invest attention into papers whose quality they perceive positively from status signals (Azoulay et al., 2013; Simcoe and Waguespack 2010; Wang and Soergel 1998). Accordingly, if after a certain time papers are ranked according to citations accrued, papers with the higher citation counts would have disproportionately substantive citations. This scenario reproduces the citation count-type pattern described above in connection with stable quality differences, and is pictured in Fig. 1A. Panel D again shows that if this pattern dominates (red curve), high citation counts understate a paper's actual influence.

On the other hand, an increasing citation count may attract disproportionately rhetorical attention. Highly cited papers may be especially useful for persuasion, a hypothesis sometimes referred to as "persuasion by name-dropping" (Frandsen and Nicolaisen 2017; White 2004). For example, sometimes citing an established expert in a particular field is used to legitimize the citer's contribution, and this function would not be fulfilled as well by referencing an obscure expert (Gilbert 1977; Mizruchi and Fein 1999). Furthermore, highly cited papers are presumably widely known to the relevant audience, making them particularly effective as commonly recognized markers in intellectual terrain (Oppenheim and Renn 1978; Small 1978). Consequently, as a paper's citation count grows, it likely becomes sufficiently attractive to an increasing number of rhetorically motivated potential citers. Eventually, the inflow of rhetorical citations may overtake the substantive. After a certain time, if papers are ranked according to citations accrued, papers with higher citation counts would have disproportionately rhetorical citations, as pictured in Fig. 1B. Panel D shows that if this pattern dominates (blue curve), high citation counts overstate papers' actual influence.

Lastly, the opposing effects of a paper's citation count proposed above may cancel out, or the arguments may simply be wrong or play only minor roles. Accordingly, if after a certain time papers are ranked according to citations accrued, their rank would have little relationship to the composition of substantive and rhetorical citations, as pictured in Fig. 1C. Panel D shows that if this pattern dominates (black line), the citation count would be a suitable proxy for influence.

Overall, this discussion suggests that both stable and time-varying characteristics of papers are associated with or even causally affect the composition of the citations they accrue. These patterns may result in potentially large biases in the extent to which citation counts measure actual influence.

3. Data and methods

We investigate the citation amount-type relationship directly using a large-scale survey of the citing authors. This approach complements the methods used in prior studies in three primary ways. First, we use direct measures of citation influence intensity - authors' reports of how much specific references influenced them. Similar data have been costly to obtain in the past. For example, the cost of measuring the quality and impact of a subset of research outputs in the UK using non-bibliometric methods has been estimated at €250 million (Else 2015). Researchers have instead examined proxies of influence, such as citations, changes in text (Gerow et al., 2018), prizes (Li et al., 2019), or recruited third parties to manually label citations for the function they appear to serve in the text (Jurgens et al., 2018; Moravcsik and Murugesan 1975; Tahamtan and Bornmann 2019; Valenzuela et al., 2015). Ultimately, however, the citing authors themselves have the best knowledge of whether their references influenced them.

Second, an experiment embedded in the survey enables a causal claim regarding the connection between a key time-varying characteristic of papers – their citation count – and their perceived quality. This link provides partial support for the argument that as papers accrue citations, they causally change how they are cited. Lastly, the sampling method ensures that the data include a wide range of disciplines from researchers around the globe. Consequently, the results provide an unusually comprehensive perspective on global research.

3.1. Sampling

The data were collected in 2018 via a personalized Qualtrics survey to randomly sampled corresponding authors of papers published in 2015 in the Web of Science database (WOS). The full survey is available in PDF format at the following address [https://github.com/MishaTeplitskiy/files/blob/main/teplitskiy_et_al_2022_citing_survey.pdf]. The year 2015 was chosen because it was the most recent year of data in our version of the database. The cited papers were published primarily in one of 15 fields, which were selected to provide both disciplinary breadth and depth (about 4500 survey solicitations per field) while staying within institutional Qualtrics usage limits. WOS attributes journals to disciplines and disciplines to six major subjects - Arts and Humanities, Clinical, Pre-clinical & Health, Engineering & Technology, Life Sciences, Physical Sciences, and Social Sciences. We sought some fields from each of these major subjects, focusing on those with large coverage by WOS and where citation-based metrics were likely to be salient. To identify the latter characteristic, for each discipline, we averaged the CiteScore of its top five journals in 2016 and then ranked disciplines according to this average, which was the primary selection criterion (Table S1). Additional criteria were that the discipline was not already used previously in an earlier pilot study and that its WOS name did not contain the word "multidisciplinary." Economics was chosen due to the authors' familiarity with it, enabling us to spot potential issues during data collection. The selected disciplines were biochemistry & molecular biology, physical chemistry, economics, endocrinology & metabolism, energy & fuels, electrical & electronic engineering, history & philosophy of science, immunology, linguistics, nanoscience & nanotechnology, oncology, pharmacology & pharmacy, applied physics, psychology, and telecommunications.

We sought a mix of relatively old and new cited papers in each discipline to enable finer-grained within-year analyses and measure how papers are cited at different points in their lifetimes. Consequently, we selected 2000, 2005, and 2010 for years of publication; at the time the citing paper was published (2015), these cited papers were about 15, 10, and 5 years old. For each of these three publication years, we ranked papers according to the number of citations they had accrued through 2015. Uncited items were included in the citation distribution. Then, from each percentile of this discipline-year specific distribution, we randomly selected five papers (cited papers) and, among the papers citing them in 2015, we selected five at random (citing papers). The choice of five citing papers per one cited paper was made to provide sufficient data for within-paper analyses. If five citing papers were not available, we randomly selected other cited papers in that percentile and repeated the procedure. The corresponding author of each citing paper for whom WOS had an email address was contacted with a personalized survey (see Appendix: Survey materials for details) and asked about two references. Given that individuals are susceptible to influence in person specific ways, and that their interpretations of survey questions may differ, we control for this by asking respondents about two papers they referenced and perform our regressions using author fixed-effects. These "within-person" models ensure that observed differences are not confounded by idiosyncratic susceptibility to "influence" or endogenous citing tendencies. The first (focal) reference was selected as above, and the second reference was chosen from the same citing paper, if possible, in the same discipline and publication year as the first. If such a second reference was not available, the restrictions were loosened until a suitable second reference could be found, or it was chosen at random. This selection procedure resulted in the set of second papers being more highly cited than the first papers on average (Figure S7). The survey finished with questions on gender and academic position ("Professor," "Associate professor," and so on).

3.2. Detailed data on citing process

To better understand the mechanisms driving the relationship, authors were also asked to identify when and how they first discovered papers they have cited, and to rate them on several dimensions of quality. Lastly, surveys were randomly assigned to display (treatment) or hide (control) the reference's citation count and percentile at the time of citing (see SI: Survey materials).

Respondents were asked to evaluate the references on several dimensions of quality (overall quality, validity, novelty, generalizability, and significance) with the question, "Rate this reference against others in the field on the following characteristics, with 50th percentile denoting the typical paper." Figure S4 shows how these dimensions were defined and the interface with which the ratings were collected.

3.3. Status signal experiment: citation count and rank

We exogenously manipulated the information respondents observed before evaluating papers. The control (85%) and treatment (15%) survey forms were identical, except the treatment form displayed the following status signal at the beginning of the survey, "Our records indicate that this paper has been cited X time(s), which ranks it in the [top/bottom] Y percentile among all papers published in the field in [year of publication]." Here, X was the paper's true citation count in Web of Science in 2015, and Y was the true percentile in the citation distribution (SI: Survey materials). Figure S2 shows the visual difference between a sample survey assigned to control vs. treatment.

If the treatment changes the perception of a paper's quality, it may have opposite effects when the citation count is relatively low vs. high. High citation counts, indicated by the count and the word "top" in "top Y percentile", constitute a positive signal that may improve perceptions and vice versa for counts in the "bottom" percentiles.

3.4. Respondents

We received completed surveys from 9380 respondents who provided data on 17,154 references. The study was approved by [Harvard University] University Committee on the Use of Human Subjects Protocol #IRB17–1320. Between February - March of 2018, we sent email solicitations to 63,049 corresponding authors of the citing papers. From this risk set, 20.2% ($n = 12{,}670$) of the recipients opened the provided link, and 15.0% ($n = 9425$) reached the last page. About 10% of emails were undeliverable. This response rate is 50–1000% percent larger than the rates obtained by other recent email-based surveys of researchers, *e.g.* (Myers et al., 2020; Radicchi et al., 2017). The respondents hailed from all over the world: 19.1% from U.S., 7.3% from China, followed by Italy, India, and Germany, with about 4–5% each. 51.5% of respondents reported employment as full, associate, or assistant professor, and 70.3% identified as male.

Response rates varied substantially across disciplines. The lowest response rate came from oncology (12.9%) and the highest (34.1%) from history and philosophy of science. The number of completed responses and response rates by discipline is displayed in Fig. 2.

Male authors, with gender inferred from names, were about 3.5% more likely to reply, while authors publishing in high impact journals were slightly less likely to reply (See SI: Nonresponse analysis for coding of gender and details of the estimation). One unit increase in impact factor was associated with a 0.73% decrease in response rate. These differences in response rates suggest that the respondents are somewhat more male and publish in slightly lower impact factor journals than the overall sample of research indexed by WOS.

In our analyses, we removed self-citations (7.3%) from the dataset because authors have been shown to evaluate their own papers much more positively than others' papers (Radicchi et al., 2017). Self-citations were self-reported via a checkbox. Although self-reported data are imperfect, bibliometric databases like WOS are likewise imperfect due to substantial identity disambiguation problems, which can lead to false positive and false negative self-citation labels. For robustness, we repeated our main analyses with self-citations based on WOS identifiers and found qualitatively similar results (available upon request).

3.5. Measuring influence and knowledge

The influence of scientific works is usually measured indirectly with citations (Zuckerman 1987). We sought to measure influence directly through self-reports. In eliciting self-reported influence, there is a trade-off between deploying a consistent but narrow definition given to respondents top-down vs a broader definition in which respondents define terms as they wish (Aksnes 2006; Radicchi et al., 2017). We opted for the former approach, given its interpretability and rarity in the literature. We measured how influential a reference was on the citing author(s) with the question, "How much did this reference influence the research choices in your paper?" which is displayed along with the answer choices in Fig. 3. Note that this definition focuses on the influence of research works on other research works, and does not address other important types of influence that research works may have, such as societal impact. We took a similar approach to elicit respondents' level of knowledge of the papers they cite (Fig. 3).

Self-reports of mental processes like influence may be problematic (Nisbett and Wilson 1977) and potentially biased by social desirability or consistency concerns. If these biases are significant, we expect that self-reported influence would be affected by the status signal experiment, with works signaled as being of higher status receiving higher influence ratings. Figure S4 and the associated Table S6 in *SI* show that the status signal did not measurably affect self-reported influence, suggesting that social desirability and self-consistency biases in self-reporting were relatively minor.

4. Results

4.1. Prevalence of rhetorical citing

First, we investigate the overall prevalence of rhetorical citing. The distribution of citations' influence intensity overall and by field is displayed in Fig. 4.

Fig. 4A shows that after removing self-citations and cases assigned to the experimental condition (14.9%), the modal citation influence level is Minor (influenced a small part of the paper, e.g., added sentence(s) to Discussion (2)). Panel B shows that only *Physical Chemistry* and *Applied Physics* showed a higher modal citation type – Moderate (influenced an important part of the paper, e.g., additional analysis). Thus, the distribution of high- and low-influence intensity citations is broadly similar across most corners of academia examined in this study. Overall, the majority (53.6%) of papers had at most "minor" influence (1 or 2) on citing authors' research choices. This is consistent with previous literature which argues that the primary goal of research articles is to position and persuade, rather than acknowledge influence (Cozzens 1989; Tahamtan and Bornmann 2018; Gilbert 1977; Moravcsik and Murugesan 1975).

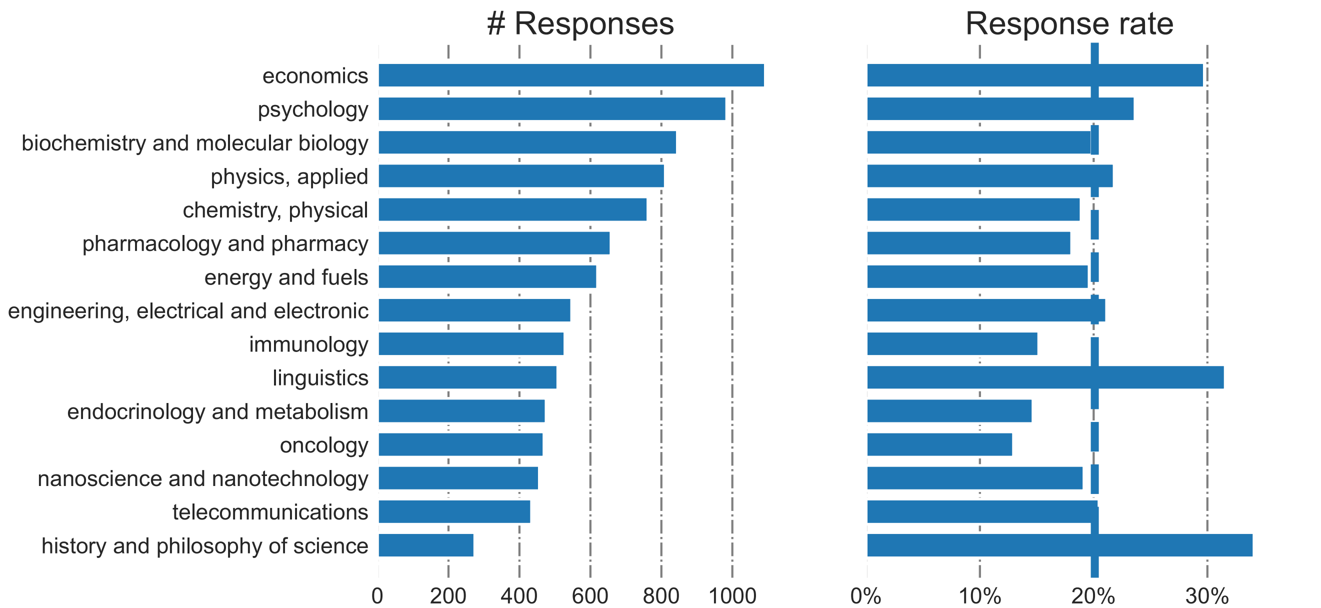

Fig. 2. Response counts and response rates by discipline. Each response, if filled out completely, provides data on two references. The dotted line shows the mean response rate.

4.2. Mean influence of marginal citation

Next, we investigate the relationship between a paper's citation count and whether its marginal citation is rhetorical or substantive. We use linear[3] mixed models with author fixed effects, which account for all stable differences between authors. This approach accounts for the possibility that the composition of authors varies significantly across the citation distribution of references. For example, authors citing lowly and highly cited works may have different standards for "influence." We use the following specification for *influence* and similarly for *knowledge*:

$$influence_{ij} = \alpha_i + \beta_0 log.cites + \beta_1 X_{ij} + \epsilon_{ij}$$

The indices *i* and *j*, enumerate authors and references, respectively. The author fixed effects denote author-specific intercepts. The set of controls X_{ij} is the indicator $added.by.coauthor_{ij}$,which equals 1 if the reference was added by a coauthor and 0 otherwise, the indicator $first.paper_{ij}$, which equals 1 if the reference appears first in the survey and 0 if second, $expertise_{ij}$ is author *i*'s self-reported expertise in paper *j*'s area, and $publication.year_j$ (normalized), the year of publication of the cited paper. log.*citations* is base-10 log of the citation count in 2015. Estimates from regressions of this form are shown in Table 1. Models 1 and 2 denote models of *influence* without and with $publication.year_j$, respectively. Models 3 and 4 denote models of *knowledge* without and with $publication.year_j$, respectively.

All models show a robust association between a paper's influence intensity, citer's knowledge of it, and log-citations. Model (2) shows that, for a given author, the influence intensity of a reference is 0.133 points on a 5-point scale higher per unit increase in log-citations. Model (4) shows that, for a given author, his or her knowledge of the reference is 0.127 points on a 5-point scale higher per unit increase in log-citations. The strength of the association was relatively unchanged when taking into account papers' age, suggesting that higher cited papers receive more substantive citations and are more closely read across their lifetimes, and not only once they become extremely highly cited.

[3] Results from logistic mixed models are qualitatively similar and available upon request.

4.3. Major influence of marginal citation

In addition to mean influence, we focus on papers that had "Major" (influenced a core part of the paper, *e.g.*, choice of theory or method) or "Very major" (motivated the entire project) influence on their readers, i.e. influence intensity of 4 or 5. These papers alter their readers' guiding theories or questions and, sometimes, motivate entirely new projects. Overall, 17.6% of references met or exceeded the major influence bar.

We define an indicator variable *major.influence* that equals 1 if the influence intensity is 4 or 5 and estimate a generalized mixed linear regression predicting it from the reference's log-citation count, *added.by.coauthor, first.paper*, and fixed effects for the citing author. The fixed effects ensure that the estimated coefficients are not affected by the highly unequal number of observations per (cited) discipline, nor by the possibly different composition of authors who use highly vs lowly cited references.

Estimates in Table 2 show a highly significant positive association between *major.influence* and log.*citations*. Additionally, we estimate a separate regression of the same form for each of 6 major subject areas of the cited papers (detailed estimates available upon request) and calculate the associated predicted probabilities of *major.influence* across a range of total citation counts. The predicted probabilities from the overall model in Table 2 and these major subject area-specific models are shown in Fig. 5.

The curve shows that the probability that a reference had a major influence is low (6.2%) for references with 1 citation but jumps to 15.6% for a reference with 1000 citations, and 20.6% for a reference with 10,000 citations. That is, citations to highly cited papers are about twice as likely, and to truly famous papers three or more times as likely, to denote major influence.

4.4. Citation count and perceptions of quality

To assess how citation counts affect perceptions of quality, we exogenously varied whether citation information was shown (treatment) or hidden (control) as respondents took the survey. The information consisted of the reference's true citation count around the time of the citation decision and ranking in the field-year-specific citation distribution and was displayed before respondents rated the reference on five dimensions of quality: overall quality, validity, significance, Reference:

[redacted]

How well do you know this paper?

- **Extremely well** (know it as well as my own work)
- **Very well** (familiar with all findings, data & methods, all limitations and critiques)
- **Well** (familiar with all findings, data & methods, some limitations)
- **Slightly well** (familiar with all findings, data & methods)
- **Not well** (only familiar with main findings)

How much did this reference influence the research choices in your paper?

- **Very major influence** (motivated the entire project)
- **Major influence** (influenced a core part of paper, e.g. choice of theory or method)
- **Moderate influence** (influenced an important part of the paper, e.g. additional analysis)
- **Minor influence** (influenced a small part of paper, e.g. added sentence(s) to Discussion)
- **Very minor influence** (paper would've been very similar without this reference)
- **Not sure**

Fig. 3. Questions used to measure influence intensity and knowledge of the cited paper.
The answer choices for the *Influence* question were converted to numbers from 1 to 5, with 1=Very minor influence and 5=Very major influence, and similarly for answers to the *Knowledge* question.

generalizability, and novelty. Figure S3 in *SI* shows the interface used in ratings and the brief description of each dimension. We expected showing citations to affect low-cited works negatively and highly cited works positively. Fig. 6A displays the associations, measured with local regression ("loess") curves, between citation percentile and perceived overall quality for control (gray curve) and treatment (red curve) observations. Loess curves are a non-parametric approach to identifying nonlinear but smooth relationships. While both curves show a positive association between citation percentile and perceived quality, the association is much stronger for the treatment papers, suggesting that the status signal exaggerated an existing quality difference.

To better quantify these patterns, we examine treatment effects on papers in the bottom vs. top half of the field-year-specific citation distribution (*SI: Status signal experiment* presents average treatment effects and additional robustness checks). The break at the median is natural because at that point the presentation of the status signal changes qualitatively, from "bottom *X*% of the citation distribution" to "top X% of the citation distribution." For this analysis we exclude observations right at the median, where the status signal presentation didn't include the words "bottom" or "top." We use the specification

$$attribute_{ij} = \beta_0 + \beta_1 above.median_{ij} + \beta_2 status.signal_i + \beta_3 above.median_{ij}{*}status.signal_i + \beta_4 X_{ij} + \in_{ij}$$

where $attribute_{ij}$ is the rating of a quality attribute by citer i of paper j, is the mean attribute for bottom-50% papers in control condition, $above.median_{ij}$ an indicator variable for top-50%, $status.signal_i$ is an indicator for treatment, X_{ij} is a set of controls of the citer and paper, and is the error. Errors are clustered by respondents, and the controls consist of the paper's discipline and publication year, the citer's gender and position, and whether the respondent or co-author added the reference. Estimates from these regressions are presented in Table 3 and displayed in Fig. 6B.

For papers below median citations, showing the status signal tends to harm quality perceptions, as indicated by negative and statistically significant coefficients of *status.signal*. The effect on perceptions of

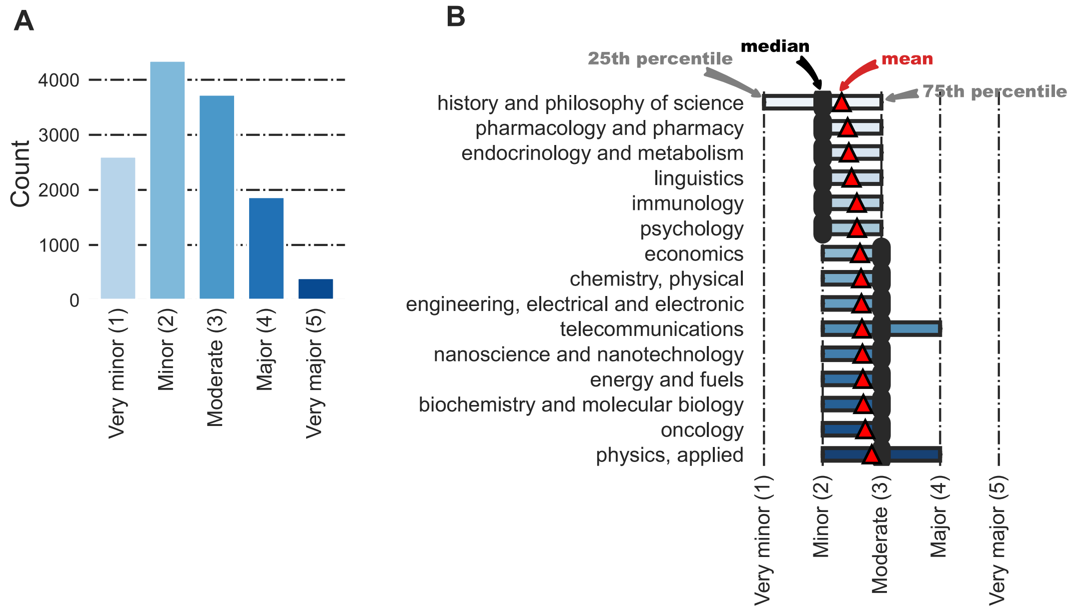

Fig. 4. **A:** Overall distribution of responses for the Influence question. **B:** Boxplot of responses by focal discipline. Each box shows the 25th percentile (left edge), the median (heavy black bar), the mean (red triangle), and the 75th percentile (right edge).

Table 1
Estimates from OLS regressions of influence (models 1 and 2) and knowledge (model 3 and 4) on log-citations, author fixed effects, and controls. Models 2 and 4 add publication year (standardized) and its interaction with log.citations. + $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$ for two-sided t-tests.

	Dependent variable:			
	Influence		Knowledge	
	(1)	(2)	(3)	(4)
log.citations	0.127***	0.125***	0.118***	0.120***
	(0.024)	(0.025)	(0.021)	(0.022)
expertise	0.372***	0.372***	0.554***	0.554***
	(0.014)	(0.014)	(0.013)	(0.013)
first.paper	−0.276***	−0.276***	−0.230***	−0.230***
	(0.019)	(0.019)	(0.016)	(0.016)
added.by.coauthor	−0.265***	−0.265***	−0.543***	−0.542***
	(0.050)	(0.050)	(0.044)	(0.044)
scale(publication.year)		0.018		
		(0.027)		(0.023)
log.citations X scale (publication.year)		−0.010		-0.025^{+}
		(0.015)		(0.013)
Citer fixed effects	Y	Y	Y	Y
Observations	12,660	12,660	12,805	12,805
R^2	0.743	0.743	0.811	0.812
Adjusted R^2	0.388	0.387	0.553	0.554
Residual Std. Error	0.830 (df = 5307)	0.830 (df = 5305)	0.736 (df = 5406)	0.736 (df = 5404)

Table 2
Estimates from a generalized linear mixed model of major.influence regressed on log-citations, controls, and citer fixed effects. + $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$ for two-sided t-tests.

	Dependent variable: major.influence
log.citations	0.343***
	(0.053)
first.paper	−0.699***
	(0.061)
added.by.coauthor	−0.726***
	(0.133)
expertise	0.815***
	(0.038)
Constant	−4.409***
	(0.247)
Citer fixed effects	Y
Observations	12,660
Log Likelihood	−5300.797
Akaike Inf. Crit.	10,613.590
Bayesian Inf. Crit.	10,658.270

quality is less precisely estimated, but is consistent with the other attributes. For papers above the median, the effect of the status signal is positive for all dimensions, although not consistently statistically significant.

Exposing citation counts produced a consistent pattern: it caused perceptions of overall quality, significance, generalizability, and novelty to fall significantly for the bottom half of papers. The effect on perceived validity matches others in direction but does not reach statistical significance. Meanwhile, exposing citations had no substantial effect on papers in the top half. If treatment effects are modeled as linear, then showing citation information harms the perceived overall quality for the bottom ~90% of papers (see *SI: Status signal experiment*). In other words, the “losers” of status signals outnumber the “winners” 9-to-1.

While treatment effects generated by displaying citations are not large (0.15–0.23 SDs), two aspects of the experiment suggest that the effects are conservative, lower-bound estimates of effects in the field. First, the respondents had already cited the papers in question, so it is likely that the citation information was not new for many. Second, other information that might plausibly signal quality was always present, such as the journal and author names.

4.5. *Citation count, discovery, and reading*

To better understand the mechanisms connecting citation count to subsequent citation influence intensity, authors were asked when and how they discovered each reference and how well they knew its content. Discovery time was measured with the question “When did you (or the co-author who added this reference) first learn about this reference?”,

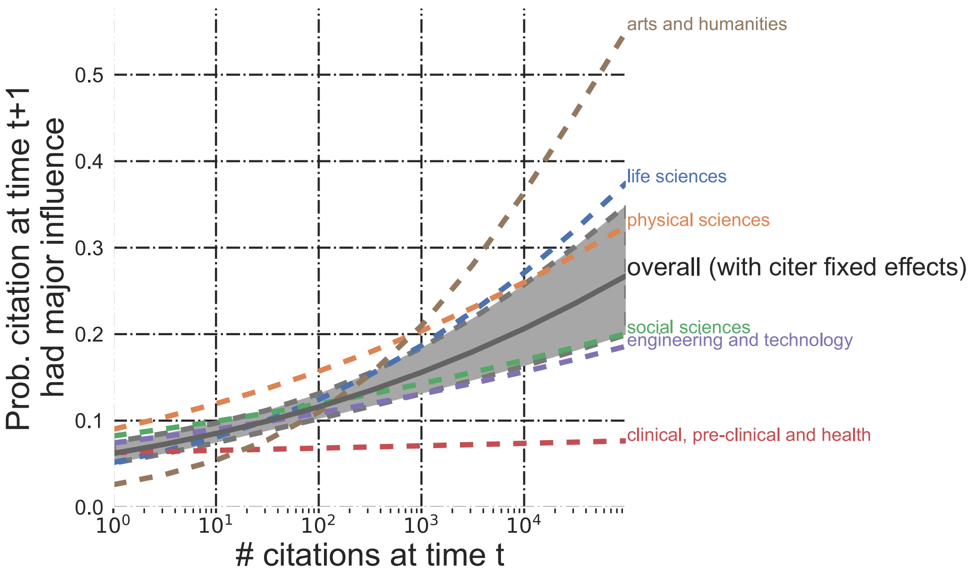

Fig. 5. Black curve with 95% CI shows predicted probability of a reference having major influence (influence ≥ 4), with fixed effects for the citer and discipline, and controls for year of publication, whether the corresponding author or coauthor added the reference, respondent's expertise in the topics of the reference, and whether it was the first paper in the survey. Colored curves show analogous probabilities for each of 6 major research areas. Marginal effects for the overall model were calculated using the ggeffects R package, using the "simulation" type of standard error, which takes into account uncertainty of the fixed and random effects.

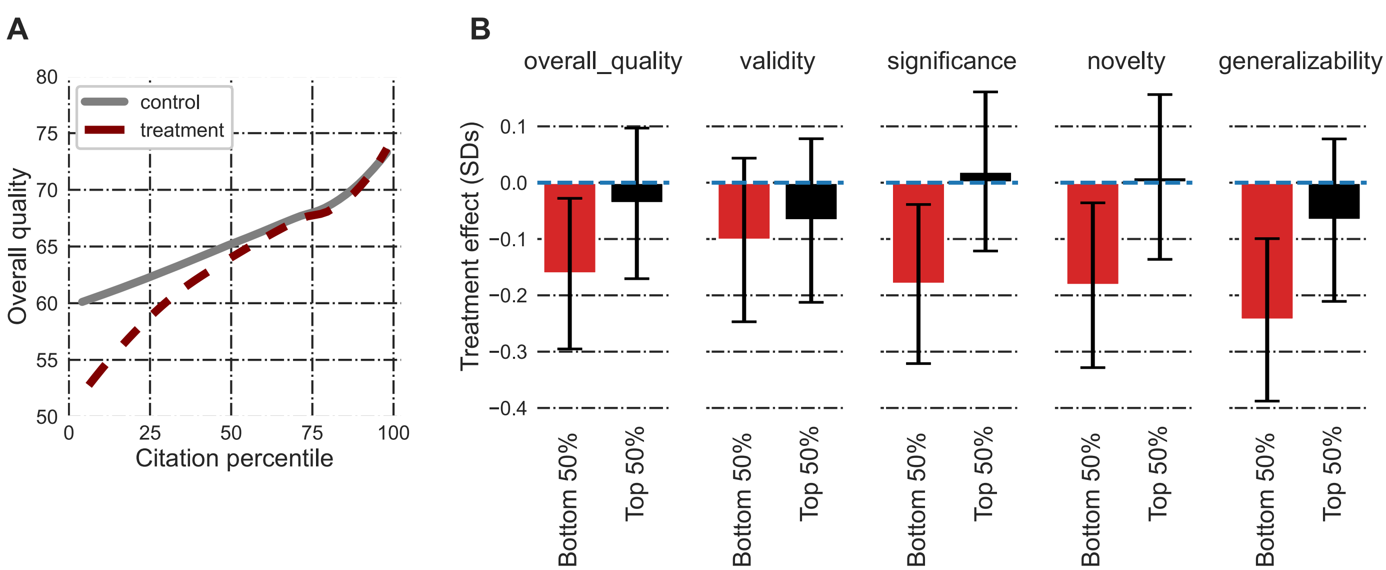

Fig. 6. **A.** Relationship between perceived "overall quality" and citation percentile of control (gray) and treatment (red) observations, measured using local regression ("loess") curves. **B.** Standardized treatment effects of status signal (coefficient of the status.signal variable in SI, Eq (3)) on five dimensions of perceived quality. Treatment effects are shown separately for references below the median in discipline-year citation distribution ("Bottom 50%") and above it ("Top 50%"), using only references published in the 15 focal fields in the 3 focal years.

Table 3
Estimates from OLS regressions of quality attribute ratings on status signal, examined above and below the median. Robust standard errors clustered at respondent level. + $p<0.1$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$ for two-sided t-tests. Constant and controls not shown.

	Dependent variable:				
	Overall_quality	Novelty	Validity	Generalizability	Significance
	(1)	(2)	(3)	(4)	(5)
above.median	4.674***	3.966***	3.146***	4.064***	4.500***
	(0.543)	(0.636)	(0.582)	(0.651)	(0.571)
status.signal	−2.899*	−3.748*	−2.058	−4.901***	−3.574**
	(1.179)	(1.479)	(1.326)	(1.456)	(1.295)
above.median X status.signal	2.379+	3.942*	0.949	3.789*	4.012**
	(1.346)	(1.647)	(1.525)	(1.659)	(1.450)
Controls	Y	Y	Y	Y	Y
Observations	7346	6822	6882	6688	7330
R^2	0.078	0.045	0.021	0.046	0.071
Adjusted R^2	0.074	0.040	0.017	0.041	0.067
Residual Std. Error	16.836 (df = 7313)	19.399 (df = 6789)	17.928 (df = 6850)	19.503 (df = 6655)	17.623 (df = 7297)

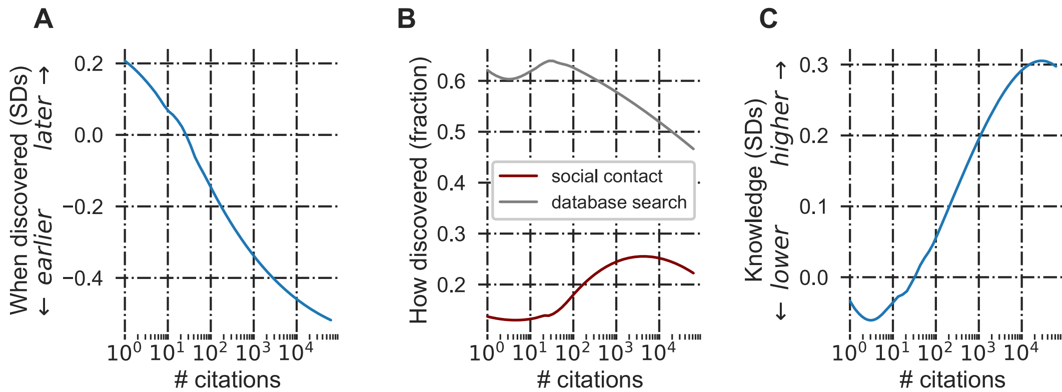

Fig. 7. **A**: When the reference was discovered (5-point scale, 1=earliest), standardized. **B:** How the reference was first discovered by the respondent (fraction of responses). Social contact consists of "recommended by colleague" and "saw in presentation." Responses without at least one of the options checked are excluded from the calculations. **C**. Respondent's knowledge of the reference's contents (1–5 point scale, 1=lowest), standardized.

with the answer choices 1=Before the project started, 2=During the project's early stages (e.g., design, data collection), 3=During the project's middle stages (e.g., analysis), 4=During the project's late stages (e. g., drafting the manuscript), 5=During the review/publication process. The method of discovery was measured with the question "How did you (or the co-author that added this reference) first learn about this reference?", with answer choices including Database search (i.e., Google Scholar), Recommended by a colleague, Saw in a conference, presentation, or class, etc.

Fig. 7 shows that highly cited papers are favored in the antecedents for influence - discovery and reading. Relative to lightly cited papers, highly cited papers are discovered earlier in the research project (Panel A), more often through social contact (Panel B). We expect papers discovered via social contact to be given more careful attention than those discovered via database search. Indeed, Panel C shows that researchers know the contents of high-status papers better than low-status ones.

While the experiment shows the role of high citation counts in generating favorable perceptions of quality, Fig. 7 shows that citations (and favorable perceptions) are in turn associated with the prerequisites of influence: early-enough discovery through meaningful channels and careful reading. Correlations between the variables displayed in Fig. 7 (*when.discovered, discovered.social.contact, discovered.database.search, knowledge*) and quality perceptions (*overall.quality*) can be found in Table S8 and qualitatively match their correlations with citation counts.

5. Discussion

These findings have several implications for measuring influence with citations. First, across most literatures, the influence intensity of a new citation is higher if the paper is already highly cited. In other words, citations to highly cited papers are, on average, more substantive in terms of the citer's engagement with the paper's contents and its influence on her research choices. Consequently, if a paper's citation count is equated with its intellectual influence, it will underestimate the actual influence of the most highly cited papers. For example, a single paper with 1000 citations is likely to be much more influential than a portfolio consisting of 100 papers with 10 citations each.

At the same time, Fig. 5 shows that in many fields, papers that had previously accrued 0–100 citations attract new citations that are of broadly similar influence intensity. This citation range is likely to include the vast majority of papers. Consequently, for most papers substantive and rhetorical citations are distributed relatively randomly, making overall citation counts a reasonable proxy of overall influence. The assumption does not hold for the most highly cited papers.

Second, the findings reveal that the trade-offs inherent in metrics like the *h*-index (Hirsch 2005) are more severe than previously appreciated. Hirsh motivated the *h*-index in part by seeking to discount "a small number of 'big hits,' which may not be representative of the individual if he or she is a coauthor with many others on those papers" (pg. 16,569). Our work does not address the co-authorship concern but does show the outsized importance of big hits. Focusing on the less cited bulk of a researcher's work may indeed be more representative of their overall output, but by ignoring the hits it becomes less informative about the person's total influence.

The findings also have several implications for inequality in science and the citing process. First, inequality in influence is usually quantified via inequality in citations, which is generally found to be large and growing (Nielsen and Andersen 2021). However, taking citation heterogeneity into account suggests that the inequality in actual influence is even larger, since citations to lightly cited papers denote systematically less influence. The findings also undermine what is sometimes referred to as the Ortega Hypothesis, that the research frontier is advanced primarily by the combined efforts of average researchers rather than a small number of elites (Cole and Cole 1972). Future tests of the hypothesis should account for the relative influence of elite papers being underestimated by their citations.

Next, we find broad support for the role of status signals, here taking the form of citation counts and ranks, in affecting how meaningfully papers are cited. Using a randomized controlled trial with a large-scale, multi-disciplinary, and global sample of researchers, we show that citations change researchers' perceptions of the quality of papers. Displaying citation counts and ranks lowered perceptions of nearly all dimensions of quality. The favorable quality perceptions generated by favorable citation counts may translate into future *substantive* citations through two key processes. Researchers discover highly cited papers more often through social connections, likely leading to closer reading, and early enough in projects to potentially change research choices. Earlier discovery and more careful reading enable such papers to have real influence, and to be cited as such. These findings corroborate those from other domains, which show that the Matthew Effect is mediated by differences in material factors: improved perceptions of quality lead audiences to allocate more resources and attention to high status actors, which help them achieve superior performance (Simcoe and Waguespack 2010; Nanda et al., 2020). Nevertheless, researchers' inferences regarding a paper's quality from citation counts are surely not the only driver of differences in discovery, reading, and influence. Given the modest effect sizes yielded by the experiment, we suspect that differences in underlying quality remain key and shape how substantively papers are cited across their lifetimes (for a similar conclusion using patents, see Higham et al., 2019).

Additionally, our work shows the utility and feasibility of using large-scale survey experiments to study perceptions of quality and influence in science. Scholars have long advocated for the use of randomized controlled trials in the science of science (Azoulay et al., 2018), but they remain rare.

This work also points to fruitful directions for future research. First, the research design relies on self-reports of influence. Although these are more direct measures than labels from third parties, identifying objective measures of influence remains a challenge. Second, our selection of disciplines to study relied on journal impact factors, leaving open the possibility that the findings do not generalize to fields where such metrics are less salient. In general, differences in citing practices across fields deserve further study. Third, we deploy a coarse-grained citation typology ("substantive" and "rhetorical") based on influence intensity. Finer-grained typologies would enable more nuanced analyses, such as the influence of positive vs. negative citations.

Fourth, we provided causal evidence that emphasizing citation status degrades the perceived quality of lightly cited papers, and that these quality perceptions are correlated with subsequent reading and citing. More work is needed to establish causal relationships across the full reading and citing pipeline. Additionally, the status signal did not materially affect the perception of papers' validity. The relative stability of perceptions of validity have been found in other work (Harris et al., 2017) and deserve further examination. An optimistic interpretation is that judgments of validity are more mechanistic. For instance, they may be determined by the community's methodological standards and are easier for researchers to make. If so, judgments of validity may be more reliable and trustworthy than judgments of other dimensions. On the other hand, a pessimistic interpretation is that researchers perceive the status signal of citation count and rank as being uninformative of works' validity. This would suggest that validity is not as strongly rewarded with citations.

Lastly, it is essential for future research to evaluate the costs and benefits of the common practice of displaying status signals like citation counts alongside papers in search and discovery contexts. Our experiment suggests that "losers" in these systems may outnumber the "winners" 9-to-1. While it is likely that displaying status signals creates more efficient search, the costs for decreased visibility of the vast majority of scholarly work should also be considered. In particular, we find that researchers' knowledge of lightly cited papers is more superficial, implying that many mistakes in perception may go uncorrected. As in other domains in which choosing to form an opinion depends on others' opinions (Bikhchandani et al., 1992) or one's initial impressions (Denrell 2005), visible citation counts can help lock in inequality in usage.

CRediT authorship contribution statement

Misha Teplitskiy: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **Eamon Duede:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Resources. **Michael Menietti:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Karim R. Lakhani:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This project benefited greatly from comments from James Evans, Ben Glick, Elena Zinchenko, and participants of the following seminars and conferences: International Conference on Computational Social Science, International Conference on Science and Technology Indicators, International Network of Analytical Sociologists Annual Meeting, University of Michigan School of Information Data Science Seminar, Innovation Growth Lab Annual Meeting, Association for Information Science and Technology Annual Meeting, National Institutes of Health Bibliometrics Symposium, and International Conference on Scientometrics & Informetrics.

Funding

This work was.ßsupported by the Harvard Business School Division of Research and Faculty Development, Laura and John Arnold Foundation, John D. and Catherine T. MacArthur Foundation, Alfred P. Sloan Foundation, and Schmidt Futures.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2022.104484.

References

Abbott, A., Cyranoski, D., Jones, N., Maher, B., Schiermeier, Q., Van Noorden, R., 2010. Metrics: do metrics matter? Nature 465 (7300), 860–862. https://doi.org/10.1038/465860a.

Aksnes, DW., 2006. Citation rates and perceptions of scientific contribution. J. Am. Soc. Inf. Sci. Technol. 57 (2), 169–185. https://doi.org/10.1002/asi.20262.

Allison, PD., Long, J.S, Krauze, TK., 1982. Cumulative advantage and inequality in science. Am. Sociol. Rev. 47 (5), 615–625. https://doi.org/10.2307/2095162.

Azoulay, P., Graff-Zivin, J., Uzzi, B., Wang, D., Williams, H., Evans, JA., Jin, G.Z., et al., 2018. Toward a more scientific science. Science 361 (6408), 1194–1197. https://doi.org/10.1126/science.aav2484.

Azoulay, P., Stuart, T., Wang, Y., 2013. Matthew: effect or fable? Manage. Sci. 60 (1), 92–109. https://doi.org/10.1287/mnsc.2013.1755.

Baldi, S., 1998. Normative versus social constructivist processes in the allocation of citations: a network-analytic model. Am. Sociol. Rev. 63 (6), 829–846. https://doi.org/10.2307/2657504.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. J. Polit. Econ. 100 (5), 992–1026. https://doi.org/10.1086/261849.

Bol, T., Vaan, M., Rijt, A., 2018. The Matthew effect in science funding. Proc. Natl. Acad. Sci. 115 (19), 4887–4890. https://doi.org/10.1073/pnas.1719557115.

Bornmann, L., Daniel, H.-D., 2008. What do citation counts measure? A review of studies on citing behavior. J. Document. 64 (1), 45–80. https://doi.org/10.1108/00220410810844150.

Brooks, TA., 1986. Evidence of complex citer motivations. J. Am. Soc. Inf. Sci. 37 (1), 34–36.

Catalini, C., Lacetera, N., Oettl, A., 2015. The incidence and role of negative citations in science. Proc. Natl. Acad. Sci., 201502280 https://doi.org/10.1073/pnas.1502280112. October.

Cole, JR., Cole, S., 1972. The Ortega Hypothesis. Science 178 (4059), 368–375. https://doi.org/10.1126/science.178.4059.368.

Cozzens, S., 1989a. What do citations count? The rhetoric-first model. Scientometrics 15 (5–6), 437–447. https://doi.org/10.1007/BF02017064.

Cronin, B., 1984. The citation process. The Role and Significance of Citations in Scientific Communication, p. 103.

Dalen, HP., Henkens, K., 2005. Signals in science - on the importance of signaling in gaining attention in science. Scientometrics 64 (2), 209–233. https://doi.org/10.1007/s11192-005-0248-5.

Denrell, J., 2005. Why most people disapprove of me: experience sampling in impression formation. Psychol. Rev. 112 (4), 951–978. https://doi.org/10.1037/0033-295X.112.4.951.

Else, H., 2015. REF 2014 Cost Almost £250 Million. Times Higher Education (THE), 2015 July 13. https://www.timeshighereducation.com/news/ref-2014-cost-250-million.

Frandsen, T.F., Nicolaisen, J., 2017. Citation behavior: a large-scale test of the persuasion by name-dropping hypothesis. J. Assoc. Inf. Sci. Technol. 68 (5), 1278–1284. https://doi.org/10.1002/asi.23746.

Gerow, A., Hu, Y., Boyd-Graber, J., Blei, DM., Evans, JA., 2018. Measuring Discursive Influence across Scholarship. Proc. Natl. Acad. Sci. 115 (13), 3308–3313. https://doi.org/10.1073/pnas.1719792115.

Gilbert, G.N, 1977. Referencing as persuasion. Soc. Stud. Sci. 7 (1), 113–122.

Harris, M., Marti, J., Watt, H., Bhatti, Y., Macinko, J., Darzi, AW., 2017. Explicit bias toward high-income-country research: a randomized, blinded, crossover experiment of english clinicians. Health Aff. 36 (11), 1997–2004. https://doi.org/10.1377/hlthaff.2017.0773.

Higham, K.W., Governale, M., Jaffe, A.B., Zülicke, U., 2019. Ex-ante measure of patent quality reveals intrinsic fitness for citation-network growth. Phys. Rev. E 99 (6), 060301.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. 102 (46), 16569–16572. https://doi.org/10.1073/pnas.0507655102.

Horbach, S., Aagaard K., and Schneider JW.. 2021. "Meta-research: how problematic citing practices distort science." MetaArXiv. 10.31222/osf.io/aqyhg.

Johnson, B., Oppenheim, C., 2007. How socially connected are citers to those that they cite? J. Document. 63 (5), 609–637. https://doi.org/10.1108/00220410710827727.

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., Jurafsky, D., 2018. Measuring the evolution of a scientific field through citation frames. Trans. Assoc. Comput. Linguist. 6 (August), 391–406. https://doi.org/10.1162/tacl_a_00028.

Krampen, G., Becker, R., Wahner, U., Montada, L., 2007. On the validity of citation counting in science evaluation: content analyses of references and citations in psychological publications. Scientometrics 71 (2), 191–202. https://doi.org/10.1007/s11192-007-1659-2.

Langfeldt, L., Reymert, I., Aksnes, D.W., 2021. The role of metrics in peer assessments. Res. Eval. 30 (1), 112–126. https://doi.org/10.1093/reseval/rvaa032.

Le Mens, G., Kovács, B., Avrahami, J., Kareev, Y., 2018. How endogenous crowd formation undermines the wisdom of the crowd in online ratings. Psychol. Sci. 29 (9), 1475–1490. https://doi.org/10.1177/0956797618775080.

Leigh Star, S., 2010. This is not a boundary object: reflections on the origin of a concept. Sci. Technol. Hum. Values 35 (5), 601–617. https://doi.org/10.1177/0162243910377624.

Li, J., Yin, Y., Fortunato, S., Wang, D., 2019. Nobel laureates are almost the same as us. Nat. Rev. Phys. 1 (5), 301–303. https://doi.org/10.1038/s42254-019-0057-z.

Liu, M., 1993. Progress in documentation the complexities of citation practice: a review of citation studies. J. Document. 49 (4), 370–408. https://doi.org/10.1108/eb026920.

Macroberts, M.H., Macroberts, BR., 1987. Testing the ortega hypothesis: facts and artifacts. Scientometrics 12 (5), 293–295. https://doi.org/10.1007/BF02016665.

MacRoberts, M.H., MacRoberts, BR., 1996. Problems of citation analysis. Scientometrics 36 (3), 435–444. https://doi.org/10.1007/BF02129604.

MacRoberts, MH., MacRoberts, BR., 1989. Problems of citation analysis: a critical review. J. Am. Soc. Inf. Sci. 40 (5), 342–349.

McKiernan, E.C., Schimanski, L.A., Nieves, C.M., Matthias, L., Niles, M.T., Alperin, J.P., 2019. "Use of the journal impact factor in academic review, promotion, and tenure evaluations." Edited by Emma Pewsey, Peter Rodgers, and Björn Brembs. Elife 8 (July), e47338. https://doi.org/10.7554/eLife.47338.

Merton, RK., 1968. The Matthew effect in science: the reward and communication systems of science are considered. Science 159 (3810), 56–63. https://doi.org/10.1126/science.159.3810.56.

Merton, RK., 1988. The Matthew effect in science, II: cumulative advantage and the symbolism of intellectual property. ISIS 79 (4), 606–623. https://doi.org/10.1086/354848.

Milard, B., 2014. The social circles behind scientific references: relationships between citing and cited authors in chemistry publications. J. Assoc. Inf. Sci. Technol. 65 (12), 2459–2468. https://doi.org/10.1002/asi.23149.

Milard, B., Tanguy, L., 2018. Citations in scientific texts: do social relations matter? J. Assoc. Inf. Sci. Technol. 69 (11), 1380–1395. https://doi.org/10.1002/asi.24061.

Mizruchi, MS., Fein, LC., 1999. The social construction of organizational knowledge: a study of the uses of coercive, mimetic, and normative isomorphism. Adm. Sci. Q. 44 (4), 653–683. https://doi.org/10.2307/2667051.

Moed, HF., Garfield, E., 2004. In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter. Scientometrics 60 (3), 295–303. https://doi.org/10.1023/B:SCIE.0000034375.39385.84.

Moravcsik, MJ., Murugesan, P., 1975. Some results on the function and quality of citations. Soc. Stud. Sci. 5 (1), 86–92. https://doi.org/10.1177/030631277500500106.

Murray, SO., Poolman, RC., 1982. Strong ties and scientific literature. Soc. Netw. 4 (3), 225–232. https://doi.org/10.1016/0378-8733(82)90023-5.

Myers, KR., Tham, W.Y., Yin, Y., Cohodes, N., Thursby, JG., Thursby, MC., Schiffer, P., Walsh, JT., Lakhani, KR., Wang, D., 2020. Unequal effects of the COVID-19 pandemic on scientists. Nat. Hum. Behav. 4 (9), 880–883. https://doi.org/10.1038/s41562-020-0921-y.

Nanda, R., Samila, S., Sorenson, O., 2020. The persistent effect of initial success: evidence from venture capital. J. Financ. Econ. 137 (1), 231–248. https://doi.org/10.1016/j.jfineco.2020.01.004.

Nicolaisen, J., 2007. Citation analysis. Ann. Rev. Inf. Sci. Technol. 41 (1), 609–641. https://doi.org/10.1002/aris.2007.1440410120.

Nielsen, M.W., Andersen, J.P., 2021. Global citation inequality is on the rise. Proc. Natl. Acad. Sci. 118 (7) https://doi.org/10.1073/pnas.2012208118.

Nisbett, RE., Wilson, TD., 1977. Telling more than we can know: verbal reports on mental processes. Psychol. Rev. 84 (3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231.

Oppenheim, C., Renn, SP., 1978. Highly cited old papers and the reasons why they continue to be cited. J. Am. Soc. Inf. Sci. 29 (5), 225–231. https://doi.org/10.1002/asi.4630290504.

Radicchi, F., Weissman, A., Bollen, J., 2017. Quantifying perceived impact of scientific publications. J. Informetr. 11 (3), 704–712. https://doi.org/10.1016/j.joi.2017.05.010.

Renear, AH., Palmer, CL., 2009. Strategic reading, ontologies, and the future of scientific publishing. Science 325 (5942), 828–832. https://doi.org/10.1126/science.1157784.

Rigney, D., 2010. The Matthew Effect: How Advantage Begets Further Advantage. Columbia University Press.

Rubin, A., Rubin, E., 2021. Systematic bias in the progress of research. J. Polit. Econ. (April) https://doi.org/10.1086/715021.

Seeber, M., Cattaneo, M., Meoli, M., Malighetti, P., 2019. Self-citations as strategic response to the use of metrics for career decisions. Res. Policy Acad. Misconduct Misrepresent. Gaming 48 (2), 478–491. https://doi.org/10.1016/j.respol.2017.12.004.

Simcoe, TS., Waguespack, DM., 2010. Status, quality, and attention: what's in a (missing) name? Manage. Sci. 57 (2), 274–290. https://doi.org/10.1287/mnsc.1100.1270.

Simmons, JP., Nelson, LD., Simonsohn, U., 2018. False-positive citations. Perspect. Psychol. Sci. 13 (2), 255–259. https://doi.org/10.1177/1745691617698146.

Small, HG., 1978. Cited documents as concept symbols. Soc. Stud. Sci. 8 (3), 327–340. https://doi.org/10.1177/030631277800800305.

Sun, M., Barry Danfa, J., Teplitskiy, M., 2021. Does double-blind peer review reduce bias? Evidence from a top computer science conference. J. Assoc. Inf. Sci. Technol. https://doi.org/10.1002/asi.24582 n/a (n/a). Accessed October 13.

Tahamtan, I., Bornmann, L., 2018. Creativity in science and the link to cited references: is the creative potential of papers reflected in their cited references? J. Informetr. 12 (3), 906–930. https://doi.org/10.1016/j.joi.2018.07.005.

Tahamtan, I., Bornmann, L., 2019. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. Scientometrics 121 (3), 1635–1684. https://doi.org/10.1007/s11192-019-03243-4.

Tenopir, C., King, DW., Christian, L., Volentine, R., 2015. Scholarly article seeking, reading, and use: a continuing evolution from print to electronic in the sciences and social sciences. Learn. Publish. 28 (2), 93–105. https://doi.org/10.1087/20150203.

Tomkins, A., Zhang, M., Heavlin, WD., 2017. Reviewer bias in single- versus double-blind peer review. Proc. Natl. Acad. Sci. U.S.A. 114 (48), 12708–12713. https://doi.org/10.1073/pnas.1707323114.

Valenzuela, M., Ha, Vu, Etzioni, O., 2015. Identifying meaningful citations. In: AAAI Workshop: Scholarly Big Data.

Wang, D., Song, C., Barabási, A.-.L., 2013. Quantifying long-term scientific impact. Science 342 (6154), 127–132. https://doi.org/10.1126/science.1237825.

Wang, P., Soergel, D., 1998. A cognitive model of document use during a research project. Study I. Document selection. J. Am. Soc. Inf. Sci. 49 (2), 115–133.

White, H., 2004. Reward, persuasion, and the Sokal Hoax: a study in citation identities. Scientometrics 60 (1), 93–120. https://doi.org/10.1023/B:SCIE.0000027313.91401.9b.

Wilhite, AW., Fong, EA., 2012. Coercive citation in academic publishing. Science 335 (6068), 542–543. https://doi.org/10.1126/science.1212540.

Zuckerman, H., 1987. Citation analysis and the complex problem of intellectual influence. Scientometrics 12 (5–6), 329–338. https://doi.org/10.1007/BF02016675.