

What Lies Within: Superscripting References to Reveal Research Trends

Eric M. Anicich

Columbia Business School

Abstract

Interpreting scholarly contributions solely on the basis of the number, and not nature, of citations is inherently flawed because contradictory as well as confirmatory findings feed into the same metric, capturing popularity at the expense of precision. I propose a citation and indexing procedure that would conveniently integrate information about research trends while imposing minimal burden on the producers and consumers of research. Under the proposed system, citations appearing in the reference list of research reports would be superscripted with letters corresponding to one of the following six categories: references to findings that are *Consistent* with the current findings, are *Replicated* by the current findings, are *Inconsistent* with the current findings, *Failed* to be replicated by the current findings, were used to build *Theory*, or were used to cite *Methodologies*. I explain how the resulting *CRIF-TM* data could be summarized and perpetually updated by an online indexing service. I provide an example to demonstrate how these superscripts could be conveniently and unobtrusively presented in the reference list of forthcoming articles. Finally, I examine the anticipated benefits, limitations, and implementation challenges of the proposed citation and indexing procedure.

Keywords

citation procedure, research practices, references

“Any publicity is good publicity” is hardly an aphorism to guide the perception of scholarly contributions. However, this is exactly how scholars have always treated published research. Indeed, the status quo citation method has long been a topic of attention and scrutiny (Abbasi, 2004; Bollen, van de Sompel, Hagberg, & Chute, 2009). However, any effort that involves interpreting scholarly contributions solely on the basis of the number, and not nature, of citations is inherently flawed because contradictory as well as confirmatory findings feed into the same metric. The number of citations simply captures popularity but does so at the expense of precision. Despite the overwhelming accumulation of new research, the scientific community steadfastly clings to this outdated and uninformative citation practice.

This uninformative citation practice contributes to the “crisis of confidence” in which the field of psychology is currently said to be embroiled (see the November 2012 *Perspectives on Psychological Science* issue dedicated to replicability for numerous views). A first step in addressing the replicability crisis is to better understand how each published article is situated within the broader academic literature. Nearly every forthcoming research

report relates to previously established findings in the literature in a variety of different ways, but with no systematic method of indexing these different relationships, there is no reasonable way to infer trends across time, articles, researchers, and fields. Systematically indexing the relationship between forthcoming findings and previously published findings would benefit every scientific discipline. In the current article, I propose a superscript referencing and indexing procedure that would conveniently integrate information about research trends and help to reduce the emergence of unfounded authority in published research (Greenberg, 2009).

Superscript Referencing and CRIF-TM Data

Citations that appear in published reports most often relate to the citing report in one of six different ways:

Corresponding Author:

Eric M. Anicich, Management Division, Columbia Business School,
3022 Broadway, Uris Hall, 7th Floor, New York, NY 10027
E-mail: eanicich16@gsb.columbia.edu

- Bendersky, C., & Hays, N. A. (2012). Status conflict in groups. *Organization Science*, 23, 323–340.^C
- Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, 42, 25–40.^I
- Crown, D. F. (2007). Effects of structurally competitive multilevel goals for an interdependent task. *Small Group Research*, 38, 265–288.^M
- Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, 85, 453–466.^M
- Groysberg, B., Polzer, J. T., & Elfenbein, H. A. (2011). Too many cooks spoil the broth: How high-status individuals decrease group effectiveness. *Organization Science*, 22(3), 722–737.^C
- Halevy, N., Chou, E. Y., Galinsky, A. D., & Murnighan, J. K. (2012). When hierarchy wins evidence from the national basketball association. *Social Psychological and Personality Science*, 3(4), 398–406.^C
- Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology*, 90, 999–1013.^T
- Leavitt, H. J. (2005). *Top down: Why hierarchies are here to stay and how to manage them more effectively*. Boston, MA: Harvard Business School Press.^C
- Markman, K. D., Lindberg, M. J., Kray, L. J., & Galinsky, A. D. (2007). Implications of counterfactual structure for creative generation and analytical problem-solving. *Personality and Social Psychology Bulletin*, 33, 312–324.^M
- Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Brain & Behavioral Sciences*, 21, 353–397.^T
- Rucker, D. D., & Galinsky, A. D. (2008). Desire to acquire: Powerlessness and compensatory consumption. *Journal of Consumer Research*, 35, 257–267.^M
- Schultheiss, O. C., Dargel, A., & Rohde, W. (2003). Implicit motives and gonadal steroid hormones: Effects of menstrual cycle phase, oral contraceptive use, and relationship status. *Hormones and Behavior*, 43, 293–301.^T
- Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review*, 10, 354–372.^C
- Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, 63, 182–196.^C
- Wade, J. B., O'Reilly, C. A. I., & Pollock, T. G. (2006). Overpaid CEOs and underpaid managers: Fairness and executive compensation. *Organization Science*, 17, 527–544.^I
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330, 686–688.^I

Fig. 1. How references would be superscripted and reported in the reference list of an article. The references are taken from the example article by Ronay, Greenaway, Anicich, and Galinsky (2012).

citations to findings that are *Consistent* (conceptually) with the current findings, are *Replicated* by the current findings, are *Inconsistent* (conceptually) with the current findings, *Failed* to be replicated by the current findings, were used to develop the *Theory* for the report, or utilized a previous *Methodology*.¹

Under the proposed system, the citations appearing in the reference list of each forthcoming report would be superscripted with letters corresponding to one of the aforementioned categories. See Figure 1 for a depiction. Summary statistics—which I refer to as *CRIF-TM* data (pronounced “crif-tum”; the hyphen separates the four directional codes from the two nondirectional codes)—could then be displayed in the form #C|#R|#I|#F|#T|#M and could be perpetually updated by an online indexing service. Figure 2 depicts how CRIF-TM data could be summarized and communicated in a succinct manner online. A few references appearing in forthcoming reports would not fall cleanly into any of the aforementioned six categories and would therefore not be superscripted.

In Table 1, I present potential interpretations of superscript referencing data for eight hypothetical reports already published in the literature. The interpretations should be treated as general guidelines. Caution should be exercised before drawing conclusions about a report

solely on the basis of the number and types of citations that a report has elicited.

The successful implementation and growth of the proposed system would require three phases. Phase 1 would be the Adoption Phase and would require fostering researcher buy-in and soliciting journals and publishers to allow superscripted references to appear in journal articles. Phase 2 would be the Migration Phase and would involve migrating CRIF-TM data to Google Scholar, Web of Science, LexisNexis, or another interdisciplinary research indexing service that would store and perpetually update CRIF-TM data and make this information available to a broad audience. If the first two phases are successful, Phase 3 would be the Scaling Phase and would involve the expansion to additional academic disciplines and retroactively superscripting the references in reports published before the introduction of the proposed system.

The proposed citation and indexing procedure would serve as an answer to recent calls to improve the integration and transparency of findings in the field. For example, Spellman (2012b) has called on researchers to work toward creating “the Decade of Putting-It-All-Together” in psychological science so as to “help us aggregate and integrate what we know” (p. 303). Specifically, Spellman’s (2012b) fourth suggestion to

Richard Ronay, Katharine Greenaway, Eric M. Anicich, and Adam D. Galinsky
 The Path to Glory Is Paved With Hierarchy: When Hierarchical Differentiation Increases
 Group Effectiveness
Psychological Science June 2012 23: 669-677, first published on May 16, 2012
 doi:10.1177/0956797611433876

CRIF-TM data: ?!?!?!?!?

[Abstract](#)

[Full Text](#)

[Full Text \(PDF\)](#)

[Request Permissions](#)

Fig. 2. Information about Ronay, Greenaway, Anicich, and Galinsky's (2012) article reproduced from *Psychological Science*'s online table of contents, with proposed placement of CRIF-TM data circled. All of this information would remain constant, with the exception of the CRIF-TM data, which would be automatically updated as more and more articles cite the target article.

implement some sort of coding system that would identify why cited work is being cited is highly relevant to this article. However, Spellman's proposal is missing an intermediate step (i.e., my proposed Phase 1) and lacks details about how journal editors would adopt and

researchers begin to embrace change to current citation practices.

Furthermore, Asendorpf et al. (2013) have outlined four ways of increasing research transparency in psychological research. Their first recommendation is to

Table 1. Potential Implications of CRIF-TM Data for Eight Hypothetical Reports Already in the Literature

Report	Consistent	Replicated	Inconsistent	Failed to replicate	Theoretical development	Methods	Potential implication for previously published report
1	++	+	+	+	+	+	Subsequent research is largely consistent with the finding(s) reported in this article; finding is likely robust.
2	+	+	++	+	+	+	Subsequent research is largely inconsistent with the finding(s) reported in this article; finding may not be robust.
3	+	++	+	+	+	+	Subsequent research has successfully replicated the finding(s) reported in this article; finding is robust.
4	+	+	+	++	+	+	Subsequent research has failed to replicate the finding(s) reported in this article; finding is not robust.
5	++	+	++	+	+	+	Subsequent research is both consistent and inconsistent with this article's findings; future researchers may seek to resolve this debate by, for example, identifying a critical moderator that may produce divergent results.
6	++	+	+	+	++	+	This article has become theoretically important but is rarely cited for other reasons; in general, this theory has been supported by subsequent findings.
7	+	+	++	+	++	+	This article has become theoretically important but is rarely cited for other reasons; in general, this theory has not been supported by subsequent findings.
8	+	+	+	+	+	++	This article has become methodologically important but is rarely cited for other reasons; it could be a pure methodology article or an empirical article that, in the context of one of the reported studies, introduced a new methodology.

Note: Column headings represent the total number of times each hypothetical report was indexed with a CRIF-TM superscript in subsequently published reports. + = small number of reports or no reports; ++ = large number of reports.

encourage researchers “to report details of the replication status of key prior studies underlying their research. Details of ‘exact’ replication studies should be reported whether they did or did not support the original study” (Asendorpf et al., 2013, p. 112). I agree in spirit with their recommendation to increase the transparency with which researchers cite relevant reports. However, their recommendation applies only to direct replication studies that are reported within the context of a forthcoming report. By providing the recommendations in the current article, I seek to bridge this gap by articulating a citation procedure that would document and index numerous types of meaningful relationships among published reports.

In the Appendix, I demonstrate the simplicity of CRIF-TM superscripting using a recent *Psychological Science* article that I coauthored (see Ronay, Greenaway, Anicich, & Galinsky, 2012). An examination of the reports cited in this article reveals a number of references that fall under the consistent, inconsistent, theoretical development, and methods categories. We did not cite any successful or failed replications.

Comparison With Legal Citation Procedures

The notion of embracing more informative citation practices is not new. The legal community has long used more comprehensive and informative citation procedures than the psychological science community in large part because the legal system in the United States relies heavily on *stare decisis*, or legal precedent, in which legal professionals are obliged, in most cases, to follow the “rules” established in previous cases. Indeed, according to Martin (2007), one of the goals of proper legal citation is to “furnish important additional information about the referenced material and its connection to the writer’s argument to assist readers in deciding whether or not to pursue the reference” (p. 3).

The nuances of legal citation are codified in *The Bluebook: A Uniform System of Citation*, a compendium of the most widely used legal citation procedures—similar to the *Publication Manual of the American Psychological Association* for psychology. According to the *Bluebook*, various citation signals are used to indicate support, contradiction, background material, and useful comparison. Whereas indexing the opinions expressed in academic legal articles may be unnecessary, documenting the treatment of previous court case rulings is critical to the legal profession. Indeed, in a legal system that is based on precedent, a citation indexing service that can provide a thorough snapshot of a case’s precedential status is critical. LexisNexis’s Shepard’s Citations do just that (see also Westlaw’s KeyCite case citator). An important part of Shepard’s Citations is the

assignment of analysis codes to indicate a court case’s complete citation history.² Negative treatments (e.g., “overruled”) are indicated with a *warning* label. Possible negative treatments (e.g., “criticized”) are indicated with a *caution* label. Positive treatments (e.g., “affirmed”) are indicated with a *positive* label. Finally, a *questioned* label is used when the validity of a citation is “questioned” by the citing reference.

Shepard’s Citations are used for legal cases and statutes, not academic publications. Therefore, the CRIF-TM superscript referencing and indexing procedure would offer a service to the field of psychology that does not currently exist in legal scholarship. Additionally, by limiting the proposed coding scheme to only a handful of important superscripts, my recommendation avoids a future mired in information overload (Spellman, 2012b).

Benefits of Superscript Referencing

The value of the proposed system is in its ability to translate what has historically only been communicated in words into quantitative metrics. In this section, I outline numerous specific benefits.

Citation summary

CRIF-TM data would serve as a quantitative footprint, quickly and efficiently revealing a more accurate “snapshot” of a report to readers than they would otherwise be able to observe by considering citation counts in the absence of qualifying information. Consider the example of a highly cited article in which a controversial finding is reported that is later refuted or debunked. Wolfe-Simon et al.’s (2011) bold finding published in *Science* that arsenic may be substituted for phosphorus to sustain bacterium growth has been strongly critiqued (see the May 27, 2011, *Science* issue for numerous comments and critiques) and empirically refuted (Reaves, Sinha, Rabinowitz, Kruglyak, & Redfield, 2012). However, as of August 8, 2014, this article had accumulated 305 Google Scholar citations. Without specific knowledge of the nature of these citations, readers may infer high quality when instead skepticism is warranted.

Just as a report’s abstract provides a concise and informative summary of the report’s findings, CRIF-TM data would provide a concise and informative summary of how the report is situated within the broader literature on the basis of the report’s relationships with subsequent reports in which it is cited. Cataloguing CRIF-TM data may also allow researchers to observe changes in citation patterns resulting from the introduction of scientific innovations, such as a new paradigm or technology. Finally, interdisciplinary scholars with less localized expertise may particularly benefit from these summary statistics,

especially as calls for open access to journals gain support (see Nosek & Bar-Anan, 2012).

Ease and flexibility of CRIF-TM data retrieval

The Shepard's Citations database on LexisNexis allows users to enter a citation to a specific case to bring up that case's entire precedential history. Indexed CRIF-TM data could eventually be accessed in a similar way by entering the article's citation or DOI number. The user would get a detailed report-level citation history that could be filtered by the type of information (e.g., superscript code) that the user is interested in. Similar searches could be conducted at the author, journal, or construct comparison level.

Streamlining the production and consumption of research

Superscript referencing would also streamline future research (e.g., literature reviews and meta-analyses) by quickly orienting readers to the nature of the references contained within a report. Take, for example, a researcher who is conducting a literature review with the aim of identifying citations to support the claim that X causes Y. After identifying a report that deals with the constructs X and Y, a quick scan of the superscripts in the reference list would reveal how many and which citations, respectively, support the desired claim. In this way, the proposed system would allow for a quick assessment of the support, or lack thereof, that exists for a particular hypothesis. Similarly, meta-analysts would be able to quickly grow their corpus of relevant articles during the information gathering stage. Overall, the roadmap that superscript referencing would produce would increase the efficiency with which readers process the contents of reports and would serve as an answer to recent calls for greater meta-analytic treatment of extant findings (Braver, Thoemmes, & Rosenthal, 2014; Stanley & Spence, 2014) while minimizing the cost imposed on authors.

Law of large numbers

In the proposed system, a vast and reliable pool of data would be generated. One implication of size is that individual coding biases would be averaged out as CRIF-TM data are aggregated across more and more independent researchers. Of course, a more general "researcher bias" may emerge in which researchers, on average, superscript references with some conscious or unconscious bias. This concern should not be discounted, and methods for dealing with such a bias are discussed later. However, the

vastness of the sample will make it statistically impossible for a single researcher to have a noticeable impact on the CRIF-TM patterns that emerge over time.

Inconsistency as opportunity

In their article titled, "The Trouble with Overconfidence," Moore and Healy (2008, p. 503) discussed the three primary forms of overconfidence—overestimation, overplacement, and overprecision—and pointed out how many "inconsistencies and problems result from treating all three varieties of overconfidence as if they were the same." The authors used the existence of these inconsistencies to motivate new theorizing and empirical investigation. For example, the authors developed new theory related to overconfidence and applied it to reconcile inconsistent findings about whether smokers overestimate (Viscusi, 1990) or underestimate (Slovic, 2001) their chance of getting lung cancer. By embracing an inconsistency in the literature as an opportunity to learn, Moore and Healy were able to advance the understanding of overconfidence in general and each of the three derivations of it in particular. Such investigations would be aided by the introduction of a more efficient citation and indexing procedure that would allow researchers to draw from a database of findings to theoretically or empirically unpack seemingly contradictory findings in the literature. In this way the "inconsistent" label should be interpreted as an invitation for researchers to dig deeper. Both Viscusi (1990) and Slovic (2001) may have been "right" because it is possible that "people have imperfect information about their own risky behavior and its consequences, but have better information about their own behavior than that of others" (Moore & Healy, 2008, p. 512); however, their findings needed to be revisited for their "rightness" to be better understood. The proposed system would document findings in need of further attention.

Minimally inconvenient

Superscript referencing would neither detract from the readability of the research nor interfere with superscripts that already appear in the body of a report, thereby imposing minimal burden on the producers and consumers of research. Furthermore, superscript referencing would occur within the scope of researchers' ordinary work flow, rather than requiring additional effort divorced from the publishing process (e.g., posting results to an external website), yet would still shed light on issues related to replicability. Over time, superscript referencing would contribute numerous data points to a vast public resource.

Scalable

The benefits of the proposed superscript referencing and indexing procedure are applicable to all scientific disciplines. Superscript referencing would allow researchers to identify trends in the literature that would not otherwise be easily perceived. Eventually, Google Scholar, Web of Science, LexisNexis, or another citation indexing service could extract and maintain CRIF-TM data from forthcoming reports, allowing nonexperts and interdisciplinary scholars to benefit from the information produced by the law of large numbers.

Advantage over the status quo

I can imagine a not too distant future in which scholars select two constructs—hierarchy and performance, for example—and one or more CRIF-TM codes—“C” and “I” for example—from a dropdown menu on a publisher’s website or online indexing service, and all published reports that support a consistent and inconsistent view of the relationship between these two constructs would appear with a click of the mouse. In the absence of superscript referencing (i.e., the world in which the field of psychology currently lives), scholars are left scouring the literature in one or more of the following ways: (a) searching for intricate combinations of relevant keywords, which, if my own experience is any indication, is an always tedious and never exhaustive effort; (b) examining already published articles on similar topics for clusters of relevant citations to support a claim one seeks to make in one’s new report, which leaves one vulnerable to the biases of the similar report’s author;³ or (c) relying on one’s own memory of the literature, an obviously flawed approach. Overall, the quality of the search results that one obtains is based too often on one’s potentially naïve ability to anticipate all relevant keyword combinations, skill in extracting citations from reports on similar topics, and confidence in and breadth of one’s own memory. Superscript referencing would reduce not only these inherent information search biases but also search time.

Limitations of Superscript Referencing

These recommendations do not come without a number of difficult challenges and limitations that the academic community would need to acknowledge and ideally address before implementing the proposed system.

References are not theory

As Sutton and Staw (1995) have pointed out, references are not theory—the same is true for CRIF-TM data. One

risk of the proposed system is that researchers may inaccurately interpret or irresponsibly use CRIF-TM data to support a specific claim. For example, researchers may (mis)use the proposed system to quickly locate a list of citations that have been coded as “consistent” with a particular relationship that they seek to investigate and then dump said citations into their article without examining the nuances of the relationships reported in the cited reports. Although CRIF-TM codes crudely capture a report’s historical treatment in the literature, relying on them for citation decisions without any analysis of the articles is strongly discouraged.

Superscript codes are categorical and assigned at the article level

A superscript code does not indicate the weight or importance of a cited report’s findings because categorical codes are assigned at the report level, whereas only certain, and in some cases secondary, findings within the cited report “earn” the code. For example, consider a report in which 90% of the theory, data, and analyses are dedicated to the relationship between X and Y (i.e., the report’s primary finding), and the remaining 10% of the theory, data, and analyses are dedicated to the relationship between X and Z (i.e., the report’s secondary finding). CRIF-TM data may obscure the true nature of the report’s findings if, for example, the report’s primary finding is highly robust, its secondary finding is not robust, and both are equally cited in subsequent reports. Such a situation may lead readers to make unwarranted inferences about the article in general without taking into account the distinction between the robustness of the article’s primary and secondary findings.

Subjectivity

The challenge of addressing coding subjectivity is perhaps the greatest threat to the viability of the proposed system.⁴ Therefore, it is paramount that a reliable coding manual is developed and circulated and that authors are trained to use it properly. Additionally, a formal mechanism should be put into place that would allow coders to seek a second, and hopefully objective, opinion for difficult coding decisions (e.g., if the coder is unsure whether a reference should be coded with a “C” or “T”).

Once there is enough stakeholder buy-in and online infrastructure to accommodate it, the scientific community may consider having objective third parties assign CRIF-TM codes to references rather than the authors themselves. One way to achieve objective CRIF-TM coding without incurring the potentially prohibitive financial cost of using paid employees is to require doctoral students, as part of

their education and training, to first receive training on proper citation procedures and then to complete a predetermined number of double-blind CRIF-TM analyses on forthcoming or already published, but unsuperscripted, articles in their area of interest. Such a system would lead to relatively more objective coding, expose doctoral students to new research relevant to their own academic interests, prepare them for academic writing, and contribute immensely to the field.

Additionally, an ongoing CRIF-TM updating option should be considered in which individuals may propose modifications to the assigned CRIF-TM codes appearing in a report. However, an objective gatekeeper would need to approve or deny the proposed modifications on the basis of some agreed on protocol.

References with multiple and seemingly contradictory superscript codes

The same reference could deserve multiple codes that seemingly contradict each other. For example, situations undoubtedly exist in which the same referenced article contains both consistent and inconsistent findings with the citing article. This may be especially true for citing or cited articles in which multiple studies are reported. Although it may seem strange to assign opposing codes to the same reference, assigning multiple codes should be encouraged when it is appropriate because multiple codes will signal the complexity of the referenced article's findings and help to avoid selective superscripting.⁵

Handling moderation

Because the answer to nearly every empirical question is "it depends," the proposed system would ideally account for moderation in some way. However, the intended "snapshot" quality of the proposed system is not conducive to a rigorous exploration of moderating variables. Rather, researchers should treat CRIF-TM data as containing potentially useful signals accompanied by some degree of noise.

Value limited to what is reported

The benefits of superscript referencing are limited to what is reported in the article. Therefore, any discussions of the "file drawer" (Rosenthal, 1979; Spellman, 2012a), questionable research practices (John, Loewenstein, & Prelec, 2012), and any other factors that may contribute to what is ultimately reported in an article are beyond the scope of this system.

Implementation and Maintenance Challenges

There are also a number of factors that would make implementing and maintaining a new citation and indexing procedure difficult.

Researcher resistance

Some researchers may feel threatened by the proposed system, believing it may challenge, in some way, the credibility of their work. This concern is not to be taken lightly. However, as I have discussed, the proposed system would do nothing more than quantitatively document and index what is already being published. The proposed procedure would merely streamline the consumption and interpretation of scientific findings.

Coordination among actors

Successfully launching and sustaining the proposed system would require a coordinated effort among numerous actors. A task force consisting of representatives from all of the major stakeholder groups could be assembled to develop an implementation plan. Bringing together all of the various stakeholder groups would not only facilitate the development of an effective implementation and maintenance plan but would also signal the legitimacy of the proposed system and open channels of communication. The status quo can only be altered if leading scholars, prominent editors, top-tier journals, established publishers, and powerful citation indexing platforms uniformly support the proposed effort.

Funding

The proposed system would undoubtedly require some level of financial investment, and securing this funding would be a significant challenge. Ideally, funding would be acquired from a nondisciplinary body so as to avoid signaling that the proposed system is intended for only one field or subfield of science. Options include seeking funding from federal scientific agencies or philanthropic funding organizations.

Encouraging adoption

Getting researchers to embrace a superscript referencing procedure would inevitably be slow. Leading scholars across disciplines would need to channel their influence to help legitimize adoption more broadly. Initially, journal editors may consider making it optional for authors to report superscripted references. At some point thereafter, journal editors may consider mandating superscript referencing.

Discussion

Reconsidering the status quo is long overdue. Jinha (2010) has estimated that there are roughly 50 million published academic articles, and this number continues to grow at a healthy rate (Larsen & von Ins, 2010). Although scientific findings have become more abundant and interrelated, the procedures by which these findings are documented and indexed have remained largely unchanged for generations. In this way, extant research can be compared with an extremely long and ever-expanding book with no table of contents—all of the accumulated knowledge is retrievable, in theory, so long as one knows where and how to look. The proposed superscripting procedure would alleviate this intractability by revealing how a particular report is situated in the larger literature.

Underlying the recommendations put forth in this article is a desire to improve the way scientific findings are reported, indexed, and subsequently cited by the scientific community. However, the potential benefits of the proposed system should be weighed against the potential limitations and implementation challenges. It is also important for researchers to distinguish between what one should and should not be licensed to infer from CRIF-TM data. CRIF-TM data are, by definition, frequency counts and should be interpreted with the same measure of caution as are other similarly crude measures. Overall,

the long-term viability and value of superscript referencing depends on researchers increasing, not decreasing, their commitment to rigorous scientific inquiry.

The work that researchers do advances science and improves society's understanding of the world. However, this pursuit is unnecessarily impeded by antiquated citation practices with limited informational value. It is time for researchers to innovate how they cite and index their research.

Example CRIF-TM superscripting for Ronay, Greenaway, Anicich, and Galinsky (2012)

Our central prediction in this article is that greater hierarchical differentiation among group members (i.e., groups with members who differed in power vs. groups with all low or all high power members) will lead to decreased interpersonal conflict and increased performance on an interdependent task (i.e., a task that requires coordination among group members). Although this article does not include any citations to successful or failed replications (i.e., “R” and “F” superscripts, respectively), recent efforts to legitimize and encourage the pursuit of preregistered replication will make the “R” and “F” codes more relevant and valuable over time.

Appendix

Table A1. Example of Superscript Referencing Decisions and Explanation

Superscript	Explanation of superscript	Excerpt from Ronay et al. (2012)
“C” for “Consistent”	References to theoretical or empirical evidence that is consistent with the citing article's	From pp. 669–670: “The central tenet of these theories is that the unequal distribution of power within groups facilitates the coordination of individuals' efforts and ultimately benefits the groups as wholes (Halevy, Chou, & Galinsky, 2011; Van Vugt, Hogan, & Kaiser, 2008) . . . Consistent with the predictions of functional theories, . . . the presence of too many high-achieving individuals within a single team had a negative effect on performance (Groysberg, Polzer, & Elfenbein, in press).”
“I” for “Inconsistent”	References to theoretical or empirical evidence that is inconsistent with the citing article's findings	From p. 669: “There is some evidence that the opposite appears to be the case, however, for humans. Research has found that inequality in groups can impair group functioning and performance. For example, more equality in members' contributions to group discussion leads to better group performance on a variety of tasks (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Similarly, wider disparities in pay increase organizational attrition (Wade, O'Reilly, & Pollock, 2006) and predict worse on-field performance in Major League Baseball (Bloom, 1999).”
“T” for “Theoretical development”	References used to develop a theoretical link that is at least one step removed from the link actually tested in the article	From p. 670: “Individual differences in testosterone predict desire for power (e.g., Schultheiss, Dargel, & Rohde, 2003) and dominance (e.g., Mazur & Booth, 1998), and high-testosterone individuals prefer being in high-power roles (Josephs, Sellers, Newman, & Mehta, 2006).” ^a

(continued)

Table A1. (continued)

Superscript	Explanation of superscript	Excerpt from Ronay et al. (2012)
"M" for "Methods"	References to methodological features of the citing article (e.g., measures, manipulations, or experimental task)	From p. 671: "To measure group productivity under conditions of high procedural interdependence, we used a modified version of Crown's (2007) letter-word-sentence game . . . To measure group productivity under conditions of low procedural interdependence, we used a creative generation task that allowed individuals to work independently within their groups (e.g., Markman, Lindberg, Kray, & Galinsky, 2007)."

"It is important to point out that these references are not specific enough to justify calling them consistent with the article's findings that hierarchical differentiation among group members leads to decreased interpersonal conflict and increased performance on a task that requires coordination among group members. Rather, these references are included in the article to develop the theoretical link between testosterone and desire for power—a relationship that, without additional context, does not directly relate to the article's findings.

Acknowledgments

I thank Andy Hafenbrack, Judy Min, and Adam Galinsky for providing outstanding feedback on earlier versions of this article.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. I acknowledge the difficulty in defining what precisely represents a successful or failed replication attempt, especially as the field continues to form consensus around the best way to conduct (Cesario, 2014; Simons, 2014; Stroebe & Strack, 2014), teach (Frank & Saxe, 2012), and reward (Koole & Lakens, 2012) replications. In my view, the necessary conditions to warrant the "R" or "F" superscript include a preregistered and approved experimental design with identical methods and analytic strategy as the original report and sufficient power. However, the exact statistical cutoff point at which a replication study succeeds or fails remains an open and important question.
2. To view a manual on the LexisNexis website that details "how to Shepardize," see http://www.lexisnexis.com/shepardscitations/printsupport/shepardize_print.pdf.
3. It is worth noting that it is virtually impossible, under the current system, for the author of one report to be fully aware of, let alone avoid falling victim to, the biases of another author.
4. There are two cases in which the superscript coding decision would be most prone to subjectivity. First is the case when determining the appropriateness of assigning an "R" (i.e., replicated) versus "C" (i.e., consistent with) code or similarly when assigning an "F" (i.e., failed to replicate) versus "I" (i.e., inconsistent with) code. I propose the following rule of thumb in these cases: If the citing author's findings are part of a preregistered replication study, then the "R" and "F" codes may be considered for the reference to which the citing author's study was designed to replicate. In short, conceptual replication attempts should always be assigned a "C" or "I" code, whereas direct replication attempts should always be assigned an "R" or "F" code. Only empirical findings may be superscripted as successful ("R") or failed ("F") replications. A second potentially difficult

distinction might occur when determining the appropriateness of assigning a "C" (i.e., consistent with) or "I" (i.e., inconsistent with) code versus a "T" (theoretical development) code. I propose the following rule of thumb in this situation: If the citing report's findings allow for a directional comparison (i.e., consistent or inconsistent) with the cited empirical evidence or theory, then the "C" or "I" codes should be considered. However, if the cited empirical evidence or theory merely acknowledges the theoretical space in which the citing report's findings are situated without being specific enough to make a directional comparison, then the "T" code should be considered.

5. Some researchers may believe that nearly all citations relate to theory development and therefore warrant "T" codes in addition to other codes potentially, but researchers are encouraged to distinguish between citations that truly develop theory for their study from those citations that only tangentially relate to one's findings. Only the former should be coded as "T" references.

References

- Abbasi, K. (2004). Let's dump impact factors. *British Medical Journal*, 329. doi:10.1136/bmj.329.7471.0-h
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108–119.
- Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal*, 42, 25–40.
- 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.
- Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9, 333–342.
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9, 40–48.
- Crown, D. F. (2007). Effects of structurally competitive multilevel goals for an interdependent task. *Small Group Research*, 38, 265–288.
- Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, 7, 600–604.

- Greenberg, S. A. (2009). How citation distortions create unfounded authority: Analysis of a citation network. *British Medical Journal*, 339. doi:10.1136/bmj.b2680
- Groysberg, B., Polzer, J. T., & Elfenbein, H. A. (2011). Too many cooks spoil the broth: How high-status individuals decrease group effectiveness. *Organization Science*, 22, 722–737.
- Halevy, N., Chou, E. Y., Galinsky, A. D., & Murnighan, J. K. (2012). When hierarchy wins evidence from the national basketball association. *Social Psychological & Personality Science*, 3, 398–406.
- Jinha, A. E. (2010). Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, 23, 258–263.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532.
- Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology*, 90, 999–1013.
- Koole, S. L., & Lakens, D. (2012). Rewarding replications A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7, 608–614.
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84, 575–603.
- Markman, K. D., Lindberg, M. J., Kray, L. J., & Galinsky, A. D. (2007). Implications of counterfactual structure for creative generation and analytical problem-solving. *Personality and Social Psychology Bulletin*, 33, 312–324.
- Martin, P. W. (2007). *Introduction to basic legal citation*. Ithaca, NY: Legal Information Institute.
- Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral & Brain Sciences*, 21, 353–397.
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115, 502–517.
- Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, 23, 217–243.
- Reaves, M. L., Sinha, S., Rabinowitz, J. D., Kruglyak, L., & Redfield, R. J. (2012). Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science*, 337, 470–473.
- Ronay, R., Greenaway, K., Anicich, E. M., & Galinsky, A. D. (2012). The path to glory is paved with hierarchy when hierarchical differentiation increases group effectiveness. *Psychological Science*, 23, 669–677.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Schultheiss, O. C., Dargel, A., & Rohde, W. (2003). Implicit motives and gonadal steroid hormones: Effects of menstrual cycle phase, oral contraceptive use, and relationship status. *Hormones and Behavior*, 43, 293–301.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9, 76–80.
- Slovic, P. (Ed.). (2001). *Smoking: Risk, perception, and policy*. Thousand Oaks, CA: Sage.
- Spellman, B. A. (2012a). Introduction to the special section: Data, data, everywhere . . . especially in my file drawer. *Perspectives on Psychological Science*, 7, 58–59.
- Spellman, B. A. (2012b). Scientific utopia . . . or too much information? Comment on Nosek and Bar-Anan. *Psychological Inquiry*, 23, 303–304.
- Stanley, D. J., & Spence, J. R. (2014). Expectations for replications are yours realistic? *Perspectives on Psychological Science*, 9, 305–318.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9, 59–71.
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, 40, 371–384.
- Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, 63, 182–196.
- Viscusi, W. K. (1990). Do smokers underestimate risks? *Journal of Political Economy*, 98, 1253–1269.
- Wade, J. B., O'Reilly, C. A. I., & Pollock, T. G. (2006). Overpaid CEOs and underpaid managers: Fairness and executive compensation. *Organization Science*, 17, 527–544.
- Wolfe-Simon, F., Blum, J. S., Kulp, T. R., Gordon, G. W., Hoeft, S. E., Pett-Ridge, J., & Oremland, R. S. (2011). A bacterium that can grow by using arsenic instead of phosphorus. *Science*, 332, 1163–1166.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330, 686–688.