



## Electronic Publication and the Narrowing of Science and Scholarship

James A. Evans, *et al.*  
*Science* **321**, 395 (2008);  
DOI: 10.1126/science.1150473

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of September 11, 2008 ):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/321/5887/395>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/321/5887/395/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/321/5887/395#related-content>

This article **cites 11 articles**, 6 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/321/5887/395#otherarticles>

This article appears in the following **subject collections**:

Sociology

<http://www.sciencemag.org/cgi/collection/sociology>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

12. A. C. Ashworth, *Anthropologie* **98**, 3 (1994).
13. All age uncertainties are reported at  $2\sigma$ . Radiocarbon ages were converted to calendar years following (27).
14. J. O. Stone, *J. Geophys. Res.* **105**, 23 (2000).
15. Materials and methods are available as supporting online material on Science Online.
16. D. C. Douglass, B. S. Singer, M. R. Kaplan, D. M. Mickelson, M. W. Caffee, *Quat. Geochron.* **1**, 43 (2006).
17. J. Mercer, *Am. J. Sci.* **266**, 91 (1968).
18. A. Gilli *et al.*, *Terra Nova* **13**, 443 (2001).
19. S. Stine, M. Stine, *Nature* **345**, 705 (1990).
20. V. Markgraf *et al.*, *Holocene* **13**, 581 (2003).
21. F. Lamy *et al.*, *Science* **304**, 1959 (2004).
22. K. J. Turner, C. J. Fogwell, R. D. McCulloch, D. E. Sugden, *Geogr. Ann.* **87A**, 363 (2005).
23. N. E. Glasser, S. Harrison, S. Ivy-Ochs, G. A. T. Duller, P. W. Kubik, *Quat. Res.* **65**, 70 (2006).
24. C. J. Heusser, *J. S. Am. Earth Sci.* **15**, 577 (2002).
25. K. D. Bennett, S. G. Haberle, S. H. Lumley, *Science* **290**, 325 (2000).
26. B. Stenni *et al.*, *Earth Planet. Sci. Lett.* **217**, 183 (2004).
27. P. J. Reimer *et al.*, *Radiocarbon* **46**, 1029 (2004).
28. We thank J. Stone for providing laboratory facilities for  $^{36}\text{Cl}$  preparation at the University of Washington. This research was supported by NSF grants to the University of

Wisconsin–Madison, Woods Hole Oceanographic Institution, and Purdue University.

### Supporting Online Material

[www.sciencemag.org/cgi/content/full/321/5887/392/DC1](http://www.sciencemag.org/cgi/content/full/321/5887/392/DC1)  
Materials and Methods  
SOM Text  
Figs. S1 to S3  
Tables S1 to S3

3 March 2008; accepted 16 June 2008  
10.1126/science.1157215

# Electronic Publication and the Narrowing of Science and Scholarship

James A. Evans

Online journals promise to serve more information to more dispersed audiences and are more efficiently searched and recalled. But because they are used differently than print—scientists and scholars tend to search electronically and follow hyperlinks rather than browse or peruse—electronically available journals may portend an ironic change for science. Using a database of 34 million articles, their citations (1945 to 2005), and online availability (1998 to 2005), I show that as more journal issues came online, the articles referenced tended to be more recent, fewer journals and articles were cited, and more of those citations were to fewer journals and articles. The forced browsing of print archives may have stretched scientists and scholars to anchor findings deeply into past and present scholarship. Searching online is more efficient and following hyperlinks quickly puts researchers in touch with prevailing opinion, but this may accelerate consensus and narrow the range of findings and ideas built upon.

Scholarship about “digital libraries” and “information technology” has focused on the superiority of the electronic provision of research. A recent Panel Report from the U.S. President’s Information Technology Advisory Committee (PITAC), “Digital Libraries: Universal Access to Human Knowledge,” captures the tone: “All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge.... In this vision, no classroom, group, or person is ever isolated from the world’s greatest knowledge resources” (1, 2). This perspective overlooks the nature of the interface between the user and the information (3). There has been little discussion of browsing/searching technology or its potential effect on science and scholarship.

Recent research into the practice of library usage measures the use of print and electronic resources with surveys, database access logs, circulation records, and reshelving counts. Despite differences in methodology, researchers agree that print use is declining as electronic use increases (4), and that general users prefer online material to print (5). These studies are also in general agreement about the three most common practices used by scientists and scholars who publish. First, most experts browse or briefly scan a small number of core journals in

print or online to build awareness of current research (6). After relevant articles are discovered online, these are often printed and perused in depth on paper (7). A second practice is to search by topic in an online article database. In recent years, the percentage of papers read as a result of browsing has dropped and been replaced by the results of online searches, especially for the most productive scientists and scholars (8). Finally, subject experts use hyperlinks in online articles to view referenced or related articles (6). Disciplinary differences exist. For example, biologists prefer to browse online, whereas medical professionals place a premium on purchasing and browsing in print. In sum, researchers peruse in print, browse in print or online (9), and search and follow citations online. These findings follow from the organization and accessibility of print and online papers. Print holdings reside either in a physical “stack” by journal and topic, arranged historically, or in a “recent publications” area. For print journals, the table of contents—its list of titles and authors—serves as the primary index. Online archives allow people to browse within journals, but they also facilitate searching the entire archive of available journals. In online interfaces where searching and browsing are both options (e.g., 3 ProQuest, Ovid, EBSCO, JSTOR, etc.), the searching option (e.g., button) is almost always placed first on the interface because logs demonstrate more frequent usage. When

searched as an undifferentiated archive of papers, titles, abstracts, and sometimes the full text can be searched by relevance and by date. Because electronic indexing is richer, experts may still browse in print, but they search online (10).

What is the effect of online availability of journal issues? It is possible that by making more research more available, online searching could conceivably broaden the work cited and lead researchers, as a collective, away from the “core” journals of their fields and to dispersed but individually relevant work. I will show, however, that even as deeper journal back issues became available online, scientists and scholars cited more recent articles; even as more total journals became available online, fewer were cited.

Citation data were drawn from Thompson Scientific’s *Science*, *Social Science*, and *Arts and Humanities* Citation Indexes, the most complete source of citation data available. Citation Index (CI) data currently include articles and associated citations from the 6000 most highly cited journals in the sciences, social sciences, and humanities going back as far as 1945, for a total of over 50 million articles. The CI flags more than 98% of its journals with from 1 to 3 of a possible 300 content codes, such as “condensed matter physics,” “omithology,” and “inorganic and nuclear chemistry.” Citation patterns were then linked with data tracking the online availability of journals from Information Today, Inc.’s *Fulltext Sources Online* (FSO).

FSO is the oldest and largest publication about electronic journal availability. Information Today began publishing FSO biannually in 1998, indicating which journals were available in which commercial electronic archives (e.g., Lexis-Nexis, EBSCO, Ovid, etc.) or if they were available freely on their own Web site, and for how many back issues. Merged together by ISSN (International Standard Serial Number), the CI and FSO data allowed me to capture how article online availability changes the use of published knowledge in subsequent research. FSO’s source distinction further allows comparison of print access with the different electronic channels through which scientists and scholars obtained articles—whether a privately maintained commercial portal or the open Internet. The combined CI-FSO data set resulted in 26,002,796

Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, IL 60615, USA. E-mail: jevans@uchicago.edu

articles whose journals came online by 2006 and a distinct 8,090,813 (in addition to the 26 million) that referenced them. Figure 1 shows the speed of the shift toward commercial and free electronic provision of articles, and how deepening backfiles have made more early science readily available in recent years.

Panel regression models were used to explore the relation between online article availability and citation activity—average historical depth of citations, number of distinct articles and journals cited, and Herfindahl concentration of citations to particular articles and journals—over time (details on methods are in the Supporting Online Material). Because studies show substantial variation in reading and research patterns by area, I used fixed-effect specifications to compare journals and subfields only to themselves over time as their online availability shifted. In this way, the pattern of citations to a journal or subfield was compared when available only in print, in print and online through a commercial archive, and online for free.

The first question was whether depth of citation—years between articles and the work they reference—is predicted by the depth of journal issues online—how many years back issues were electronically available during the previous year when scientists presumably drafted them into their papers. For subfields, this was calculated as years from the first journal's availability. These data were collected in publication windows of 20 years, and so only data from 1965—20 years after the beginning of the data set—were used. For the entire data set, citations pointed to articles published an average of 5.6 years previously (table S1). The average number of years journal articles were available online is only 1.85 (the data go back to 1945), but with a standard deviation of 5 years and a

maximum of more than 60 years. Analysis was performed by citation year and within journal or subfield. The standard ordinary least squares (OLS) method for linear regression was used in generating all the results to be described.

All regression models contained variables used to account and statistically control for alternative explanations of why citations might refer to more recent articles. A sequence of integers from 1 to 40, corresponding to citation years 1965 through 2005, was included to account for a general trend of increasing citations over time (the estimates for this variable were always positive and statistically significant,  $P < 0.001$ ). Average number of pages and average number of references in citing articles were both included to account for the possibility that citations are more recent because articles are shorter with fewer references and the earliest ones have been disproportionately “censored” by publishers (estimates for pages were positive but not always significant; those for references were always positive and significant,  $P < 0.001$ : longer articles with more references referred to earlier work). A measure of the average age of title words was also included in the models to account for the possibility that in recent years, research has concerned more recent concepts or recently discovered (or invented) phenomena. This was calculated by taking the age of each title word within the relevant publication window for the analysis (e.g., prior 20 years) and then multiplying it by a weight for each word  $i$  in title  $j$

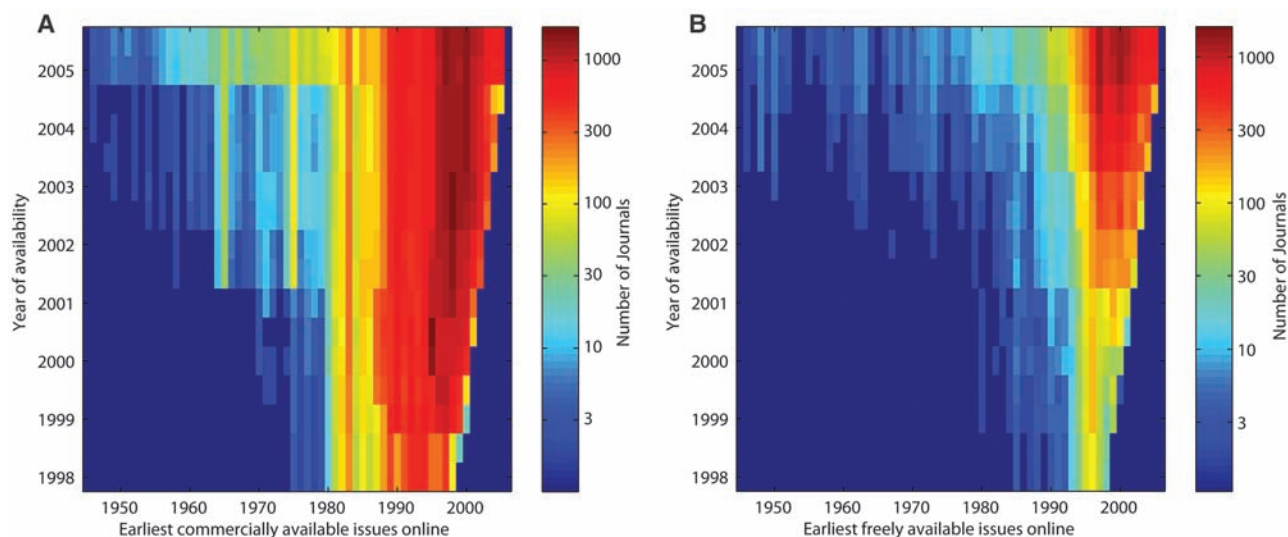
equivalent to  $\sum_{i=1}^k (1 + \ln(tf_{ij})) \times \ln\left(\frac{N}{df_i}\right)$  where

$tf_{ij}$  equals the frequency of term  $i$  in title  $j$  and  $df_i$  equals the number of articles in a given year that contain term  $i$  out of the total number

of annual articles  $N(I)$ . This approach highly weights distinguishing title terms (e.g., *buckyballs*, *microRNA*) and gives lesser weight to broad area terms (e.g., *gene*, *ocean*) and virtually no weight to universal words (e.g., *and*, *the*). Regression coefficients for the title age measures were always positive and significant ( $P < 0.0001$ ), indicating that titles with older terms referenced earlier articles. Each model also contained a constant with a significant negative estimate.

The graphs in Fig. 2 trace the influence of online access, estimated from the entire sample of articles, and illustrated for journals and subfields with the mean number of citations. Figure 2A shows the simultaneous effect of commercial and free online availability on the average age of citations. Consider a journal whose articles reference prior work that is, on average, 5.6 years old—the sample mean. If that journal's issues become available online for an additional 15 years, both commercially and for free, the average age of references would decrease to less than 4.5 years, falling by 0.088 years for each new online year available. The within-subfield models followed the same pattern, although confidence intervals were wider (tables S2 to S4).

To determine the effect of online availability on the amount of distinct research cited, I explored the relation between the distinct number of articles and journals cited in a given citation year by depth of online availability. The number of distinct articles and journals was calculated over a 20-year window, as in the previous analysis. For the average journal, 632 articles were cited each year, but this ranges widely. Because citation values are discrete and because high values concentrate within a few core journals but vary widely among the others, I modeled its relation with online availability by means of neg-



**Fig. 1.** Distribution of online journal availability in ISI-F50 data through (A) commercial subscription and (B) free through journal Web site. “Hot” regions of the graph correspond to journal issues just a few years behind the years in which they are available online, e.g., in 2003, more journals

were commercially and freely available from 1999—about 1000 and 500, respectively—than from any other year. The figure highlights how journal issues increasingly came online from the 1940s, '50s, and '60s in 2004 and 2005.



ative binomial models (12). The negative binomial is a generalization of the Poisson model that allows for an additional source of variance above that due to pure sampling error. A fixed-effects specification of this model refers not to the coefficient estimates but to the “dispersion parameter,” forcing the estimated variance of citations to be the same within journals or subfields, but allowing it to take on any value across them. These models were estimated with the maximum likelihood method and produced coefficient estimates that, when exponentiated, can be interpreted as the ratio of (i) the number of distinct articles cited after a 1-year increase in the electronic provision of journals over (ii) the number of articles cited without an online increase. One can subtract 1 from these ratios and

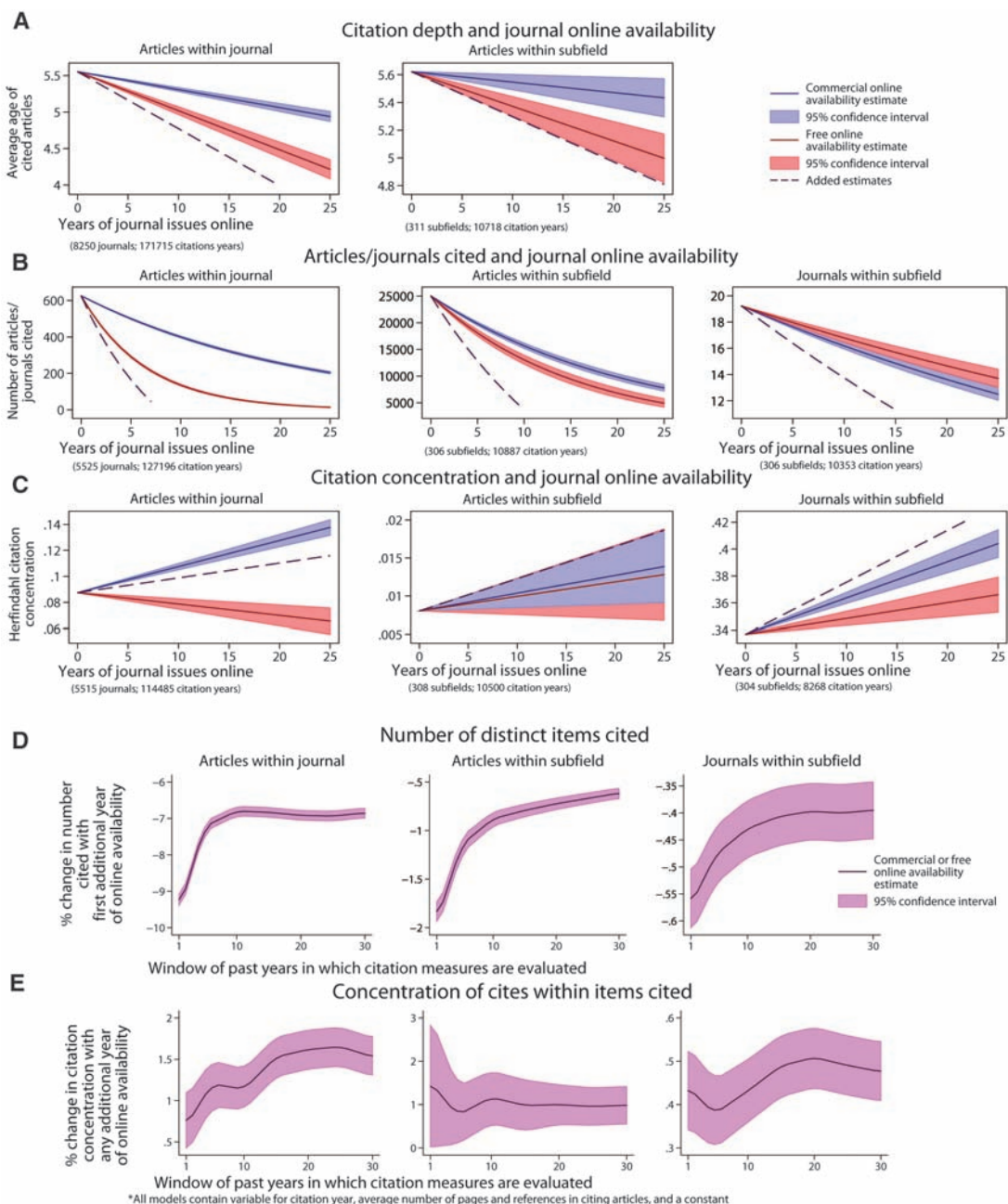
multiply by 100 to obtain the percentage change of a 1-year increase in online availability on the number of distinct items cited. All models contained measures that statistically control for citation year, average number of pages, and references in citing articles.

In each subsequent year from 1965 to 2005, more distinct articles were cited from journals and subfields. The pool of published science is growing, and more of it is archived in the CI each year. Online availability, however, has not driven this trend. Figure 2B illustrates the simultaneous effect of free and online availability on the number of distinct articles cited in journals, and the number of distinct articles and journals cited in subfields. The panels portray these effects for a hypothetical journal and sub-

field receiving the sample mean of citations. With five additional years of free and commercial online availability, the number of distinct articles cited within journal would drop from 600 to 200; the number of articles cited within subfields would drop from 25,000 to 15,000; and the number of journals cited within subfields would drop from 19 to 16. This suggests that online availability may have reduced the number of distinct articles and journals cited below what it would have been had journals not gone online. Provision of one additional year of issues online for free associates with 14% fewer distinct articles cited.

Fewer distinct articles and journals were cited soon after they went online. Although this influenced the overall concentration of article

**Fig. 2.** Estimated influence of commercial and free online article availability (in years of journal issues available online) on (A) mean age of citations (based on OLS regression coefficients); (B) distinct number of articles and journals cited (based on exponentiated maximum likelihood negative binomial regression coefficients); and (C) Herfindahl concentration of citations within particular articles and journals (based on OLS regression coefficients). Each of these relations is illustrated relative to the sample mean of citation age, number, and concentration; each relation illustrated represents an underlying model that accounts for citation year, number of pages, and number of references in citing articles; the underlying citation age model also accounts for the mean weighted age of weighted title words in citing articles. Estimated percentage change, given one additional year of online availability, for (D) number of distinct articles and journals cited and (E) Herfindahl concentration within those citations, when enlarging the window in which citation measures are evaluated, from 1 to 30 years—1975 to 2005.



Downloaded from www.sciencemag.org on September 11, 2008

citations in science, it did not fully determine it. Citations may be spread more evenly over fewer articles to more broadly disperse scientific attention. To assess the degree to which online provision influences the concentration of citations to just a few articles (and journals), I computed a Herfindahl index, where  $\sum_{j=1}^n (s_j^2)$  represents the percentage of citations  $s$  to each article  $j$ , squared and summed across journal or subfield  $i$  within the 20-year time window examined. A concentration of 1 indicates that every citation to journal  $i$  in a given year is to a single article; a concentration just less than 1 suggests a high proportion of citations pointing to just a few articles; and a concentration approaching zero implies that citations reach out evenly to a large number of articles. Herfindahl concentrations of articles cited in journals ranged from 0.0000933 to 1 in this sample, with an average of 0.088 and a wide standard deviation of 0.195. Where no articles were cited, no concentrations could be computed. Regression models were used to examine whether citation concentration to articles from the last 20 years could be attributed to depth of online availability. As in previous models, these were estimated for articles within journals and for articles and journals within subfields, by means of both commercial and free electronic provision. Citation concentrations are approximately normally distributed and the models were estimated with OLS.

Figure 2C illustrates the concurrent influence of commercial and free online provision on the concentration of citations to particular articles and journals. The left panel shows that the number of years of commercial availability appears to significantly increase concentration of citations to fewer articles within a journal. If an additional 10 years of journal issues were to go online via any commercial source, the model predicts that its citation concentration would rise from 0.088 to 0.105, an increase of nearly 20%. Free electronic availability had a slight negative effect on the concentration of articles cited within journals, but it had a marginally positive effect on the concentration of articles cited within subfields (middle panel) and appeared to substantially drive up the concentration of citations to central journals within subfields (right panel). Commercial provision had a consistent positive effect on citation concentration in both articles and journals. The collective similarity between commercial and free access for all models discussed suggests that online access—whatever its source—reshapes knowledge discovery and use in the same way. For all models, similar results were obtained when journals' presence in multiple (e.g., one, two, and three or more) commercial archives was accounted for and modeled simultaneously.

Although 20 years is not an unreasonable window of time within which to examine the effect of online availability on citations, it does not capture the trend of the effect. For example,

one can imagine that online provision increases the distinct number of articles cited and decreases the citation concentration for recent articles, but hastens convergence to canonical classics in the more distant past. To explore this possibility, I performed the same analyses but calculated variables with expanding windows ranging from the last year to the last 30 years. To keep samples comparable, I estimated all models on data from 1975 (1945 plus a 30-year window) to 2005, and so the 20-year window coefficients do not correspond perfectly to the effects illustrated earlier. Estimated percentage changes in the number of articles and journals cited and the Herfindahl citation concentration within those citations were calculated as associated with a 1-year extension of online availability. These estimates and their corresponding 95% confidence intervals are graphed in Fig. 2, D and E. Increased online provision in the preceding year was associated with a decrease in the number of distinct articles cited within journals and articles and journals cited within subfields most in recent years (Fig. 2D). A 1-year change in online availability corresponded to a 9% drop in articles cited in the last year, but only a 7% drop in articles cited in the past 20 and 30 years. The pattern was the same for articles and journals within subfields (tables S2 to S4). The citation window's effect on citation concentration was not so consistent (Fig. 2E). Nevertheless, in the case of article concentrations within subfields, the Herfindahl concentration increase was highest—1.5% per year of online availability—when calculated for references to only the last year's articles.

The models presented are limited in a number of ways. For example, journals such as *Science* use Supporting Online Material for "Materials and Methods," which frequently include references not indexed by the CI. It is theoretically possible, though unlikely, that these references are to earlier or more diverse articles. Moreover, by studying only conventional journals, this study fails to capture newer scientific media like science blogs, wikis, and online outlets exploring alternative models of peer review. These new media almost undoubtedly link to extremely recent scientific developments—often through ephemeral Web links (13)—but they may also point to more diverse materials.

Collectively, the models presented illustrate that as journal archives came online, either through commercial vendors or freely, citation patterns shifted. As deeper backfiles became available, more recent articles were referenced; as more articles became available, fewer were cited and citations became more concentrated within fewer articles. These changes likely mean that the shift from browsing in print to searching online facilitates avoidance of older and less relevant literature. Moreover, hyperlinking through an online archive puts experts in touch with consensus about what is the most important prior

work—what work is broadly discussed and referenced. With both strategies, experts online bypass many of the marginally related articles that print researchers skim. If online researchers can more easily find prevailing opinion, they are more likely to follow it, leading to more citations referencing fewer articles. Research on the extreme inequality of Internet hyperlinks (14), scientific citations (15, 16), and other forms of "preferential attachment" (17, 18) suggests that near-random differences in quality amplify when agents become aware of each other's choices. Agents view others' choices as relevant information—a signal of quality—and factor them into their own reading and citation selections. By enabling scientists to quickly reach and converge with prevailing opinion, electronic journals hasten scientific consensus. But haste may cost more than the subscription to an online archive: Findings and ideas that do not become consensus quickly will be forgotten quickly.

This research ironically intimates that one of the chief values of print library research is poor indexing. Poor indexing—indexing by titles and authors, primarily within core journals—likely had unintended consequences that assisted the integration of science and scholarship. By drawing researchers through unrelated articles, print browsing and perusal may have facilitated broader comparisons and led researchers into the past. Modern graduate education parallels this shift in publication—shorter in years, more specialized in scope, culminating less frequently in a true dissertation than an album of articles (19).

The move to online science appears to represent one more step on the path initiated by the much earlier shift from the contextualized monograph, like Newton's *Principia* (20) or Darwin's *Origin of Species* (21), to the modern research article. The *Principia* and *Origin*, each produced over the course of more than a decade, not only were engaged in current debates, but wove their propositions into conversation with astronomers, geometers, and naturalists from centuries past. As 21st-century scientists and scholars use online searching and hyperlinking to frame and publish their arguments more efficiently, they weave them into a more focused—and more narrow—past and present.

#### References and Notes

1. R. Reddy *et al.*, "Digital Libraries: Universal Access to Human Knowledge" (President's Information Technology Advisory Committee, Panel on Digital Libraries, 2001); [www.nitrd.gov/pubs/pitac/pitac-dl-9feb01.pdf](http://www.nitrd.gov/pubs/pitac/pitac-dl-9feb01.pdf).
2. The report (2) qualifies the vision of universal access, but only by admitting that "more 'quality' digital contents" must be made available and better IT infrastructure must deliver them.
3. M. McLuhan, *Understanding Media* (McGraw-Hill, New York, 1964), chap. 1.
4. S. Black, *Libr. Resour. Tech. Serv.* **49**, 19 (2005).
5. S. L. De Groot, J. L. Dorsch, *J. Med. Libr. Assoc.* **91**, 231 (2003).
6. C. Tenopir, B. Hitchcock, S. A. Pillow, "Use and Users of Electronic Library Resources: An Overview and Analysis of

- Recent Research Studies" (Council on Library and Information Resources, Washington, DC, 2003).
- A. Friedlander, "Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc." (Council on Library and Information Resources, Washington, DC, 2002); [www.clir.org/pubs/reports/pub110/contents.html](http://www.clir.org/pubs/reports/pub110/contents.html).
  - P. Boyce, D. W. King, C. Montgomery, C. Tenopir, *Ser. Libr.* **46**, 121 (2004).
  - C. Tenopir, D. W. King, A. Bush, *J. Med. Libr. Assoc.* **92**, 233 (2004).
  - C. Shirky, "Ontology is Overrated: Categories, Links and Tags" (Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community, Open Source, 2005); [www.shirky.com/writings/ontology\\_overrated.html](http://www.shirky.com/writings/ontology_overrated.html).
  - C. Manning, H. Schütz, *Foundations of Natural Language Processing* (MIT Press, Cambridge, MA, 1999).
  - J. Hausman, B. H. Hall, Z. Griliches, *Econometrica* **52**, 909 (1984).
  - R. P. Dellavalle *et al.*, *Science* **302**, 787 (2003).
  - A. L. Barabási, R. Albert, *Science* **286**, 509 (1999).
  - R. K. Merton, *Science* **159**, 56 (1968).
  - D. J. de Solla Price, *Science* **149**, 510 (1965).
  - H. A. Simon, *Biometrika* **42**, 425 (1955).
  - M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854 (2006).
  - J. Berger, "Exploring ways to shorten the ascent to a Ph.D.," *New York Times*, 3 October 2007; [www.nytimes.com/2007/10/03/education/03education.html](http://www.nytimes.com/2007/10/03/education/03education.html).
  - I. Newton, *Principia* (Macmillan, New York, ed. 4, 1883) (first published in 1687).
  - C. Darwin, *The Origin of Species* (D. Appleton, New York, 1867) (first published in 1859).
  - I gratefully acknowledge research support from NSF grant 0242971, Science Citation Index data from Thompson Scientific, Inc., and *Fulltext Sources Online* data from Information Today, Inc. I also thank J. Reimer for helpful discussion and insight.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/321/5887/395/DC1](http://www.sciencemag.org/cgi/content/full/321/5887/395/DC1)  
Methods  
Tables S1 to S4  
References

13 September 2007; accepted 9 June 2008  
10.1126/science.1150473

# The Evolution and Distribution of Species Body Size

Aaron Clauset<sup>1\*</sup> and Douglas H. Erwin<sup>1,2</sup>

The distribution of species body size within taxonomic groups exhibits a heavy right tail extending over many orders of magnitude, where most species are much larger than the smallest species. We provide a simple model of cladogenetic diffusion over evolutionary time that omits explicit mechanisms for interspecific competition and other microevolutionary processes, yet fully explains the shape of this distribution. We estimate the model's parameters from fossil data and find that it robustly reproduces the distribution of 4002 mammal species from the late Quaternary. The observed fit suggests that the asymmetric distribution arises from a fundamental trade-off between the short-term selective advantages (Cope's rule) and long-term selective risks of increased species body size in the presence of a taxon-specific lower limit on body size.

Most taxonomic groups show a common distribution of species body size (1–3), with a single prominent mode relatively near but not at the smallest species size (4) and a smooth but heavy right tail (often described as a right skew on a log-size scale) extending for several orders of magnitude (e.g., Fig. 1). This distribution is naturally related to a wide variety of other species characteristics with which body size correlates, including habitat, life history, life span (5), metabolism (6), and extinction risk (7). A greater understanding of the underlying constraints on, and long-term trends in, body size evolution may provide information for conservation efforts (8) and insight about interactions between ecological and macroevolutionary processes (9).

Studies of body-size distributions have suggested that the prominent mode may be indicative of a taxon-specific energetically optimal body size (10, 11), which is supported by microevolutionary studies of insular species (12). However, evidence for Cope's rule (1, 13, 14)—the observation that species tend to be larger than their ancestors—and the fact that most species are not close to their group's predicted optimal size [among other reasons (15)] suggest that this

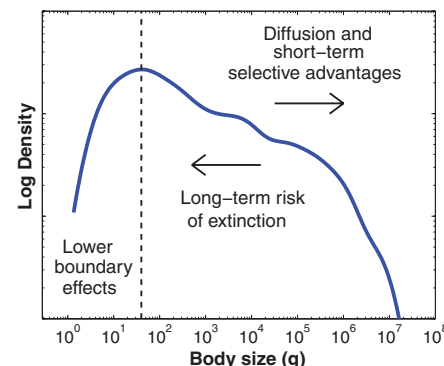
theory may be flawed. Alternatively, species body sizes may diffuse over evolutionary time. If so, Cope's rule alone could cause size distributions to exhibit heavy right tails (1), although size-dependent speciation or extinction rates (2, 9, 16) or size-neutral diffusion near a taxon-specific lower limit on body size (17) could also produce a similar shape. Furthermore, different mechanisms may drive body-size evolution on spatial and temporal scales (3), and the importance of interspecific competition to the macroevolutionary dynamics of species body size is not known.

We developed a generalized diffusion model of species body-size evolution, in which the size distribution is the product of three macroevolutionary processes (Fig. 1). We combine these processes, each of which has been independently studied (1, 2, 17, 18), in a single quantitative framework, estimate its parameters from fossil data on extinct terrestrial mammals from before the late Quaternary (19, 20), and determine whether this model, or simpler variants, can reproduce the sizes of the 4002 known extant and extinct terrestrial mammal species from the late Quaternary (Recent species) (21, 22).

This model makes three assumptions: (i) Species size varies over evolutionary time as a cladogenetic multiplicative diffusion process (1, 17); the size of a descendant species  $x_D$  is the product of a stochastic growth factor  $\lambda$  and its ancestor's size  $x_A$ , that is,  $x_D = \lambda x_A$ . For each speciation event, a new  $\lambda$  is drawn from the dis-

tribution  $F(\lambda)$ , which models the total influence on species size changes from all directions. A bias toward larger sizes (Cope's rule) appears as a positive average log-change to size ( $\log \lambda > 0$ ), and may depend on the ancestor's size. (ii) Species body size is restricted by a taxon-specific lower limit  $x_{\min}$  (6, 23), which we model by requiring that  $F(\lambda < x_{\min}/x_A) = 0$ , that is, the largest possible decrease in size for a particular speciation event is  $\lambda = x_{\min}/x_A$ . In our computer simulations, time proceeds in discrete steps. At each step, exactly one new species is produced, which is the descendant of a randomly selected species. (iii) Every species independently becomes extinct with probability  $p_e(x)$ , which increases monotonically with size. A schematic of the model is shown in Fig. 2A. [For technical details, see (24).]

To make this model appropriately realistic, we estimated the form of each process from fossil data. The lower limit on mammalian body size is near 2 g, close to the size of both the Etruscan shrew (*Suncus etruscus*) and the bumblebee bat (*Craseonycteris thonglongyai*). Fossil evidence suggests that this limit has existed since at least the Cretaceous-Tertiary boundary (19, 20, 25). Further, a limit in this vicinity is supported by both



**Fig. 1.** Smoothed species body size distribution of 4002 Recent terrestrial mammals [data from (21)], showing the three macroevolutionary processes that shape the relative abundances of different sizes. The left tail of the distribution is created by diffusion in the vicinity of a taxon-specific lower limit near 2 g, whereas the long right tail is produced by the interaction of diffusion over evolutionary time (including trends like Cope's rule) and the long-term risk of extinction from increased body size.

<sup>1</sup>Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA. <sup>2</sup>Department of Paleobiology, MRC-121, National Museum of Natural History, Post Office Box 37012, Washington, DC 20013-012, USA.

\*To whom correspondence should be addressed. E-mail: [aaronc@santafe.edu](mailto:aaronc@santafe.edu)